Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Classificação automática de emoções utilizando imagens faciais: Uma abordagem a partir da Magnificação de Vídeo Euleriana e Redes Neurais Artificiais

Vitor Quaresma Silveira de Hollanda Ramos

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Orientador
Prof. Dr. Flávio de Barros Vidal

Brasília
2015

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Curso de Engenharia da Computação

Coordenador: Prof. Dr. Ricardo Zelenovsky

Banca examinadora composta por:

Prof. Dr. Flávio de Barros Vidal (Orientador) — CIC/UnB
Prof. Dr. Alexandre Ricardo Soares Romariz — ENE/UnB
Prof. Dr. Alexandre Zaghetto — CIC/UnB

CIP — Catalogação Internacional na Publicação

Ramos, Vitor Quaresma Silveira de Hollanda.

Classificação automática de emoções utilizando imagens faciais: Uma abordagem a partir da Magnificação de Vídeo Euleriana e Redes Neurais Artificiais / Vitor Quaresma Silveira de Hollanda Ramos. Brasília : UnB, 2015.
64 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2015.

1. Detecção de emoções, 2. Análise facial, 3. Magnificação de Vídeo Euleriana

CDU 004

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília–DF — Brasil

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Classificação automática de emoções utilizando imagens faciais: Uma abordagem a partir da Magnificação de Vídeo Euleriana e Redes Neurais Artificiais

Vitor Quaresma Silveira de Hollanda Ramos

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Prof. Dr. Flávio de Barros Vidal (Orientador)
CIC/UnB

Prof. Dr. Alexandre Ricardo Soares Romariz
ENE/UnB

Prof. Dr. Alexandre Zaghetto
CIC/UnB

Prof. Dr. Ricardo Zelenovsky
Coordenador do Curso de Engenharia da Computação

Brasília, 11 de dezembro de 2015

Dedicatória

Dedico este trabalho às pessoas que foram pacientes, me apoiaram e sempre acreditaram em mim na realização deste trabalho. Dedico este aos meus amigos e minha família pela felicidade que me proporcionam e dedico também as pessoas que despertam em mim sentimentos que não podem ser explicados e que nenhum classificador consegue identificar, Diva, Daniela, Laura, Gabriel, Manoel, Cláudia e Gisela.

Agradecimentos

Um agradecimento à Universidade de Brasília por me fornecer o conhecimento necessário para realizar este trabalho. Um agradecimento ao pessoal do LISA e aos meus colegas de curso pelo apoio, pela paciência e pelos momentos de descontração. Agradecimento ao meu colega Mateus Mendelson pelo apoio e ao professor Flávio Vidal por ter me orientado e não ter desisitido da ideia.

Resumo

Neste trabalho, uma abordagem inicial foi proposta para desenvolver um classificador automático de emoções usando Magnificação de Vídeo Euleriana e Redes Neurais Artificais. A abordagem proposta cria um vetor de descritores dos vídeos processados com Magnificação Euleriana e usa classificação pela Rede Neural Artificial para alcançar uma taxa aceitável de acurácia para emoções combinadas e isoladas. Para validar a eficiência do classificador de emoções, foram geradas matrizes de confusão com a emoção classificada pelo classificador e a emoção real do vídeo, sendo realizados testes com diversos conjuntos de emoções. O classificador conseguiu classificar de forma eficiente conjuntos de até 3 classes de emoções, se confundindo com mais emoções.

Palavras-chave: Detecção de emoções, Análise facial, Magnificação de Vídeo Euleriana

Abstract

In this work, an initial approach was proposed to develop an automatic emotion classification using Eulerian Video Magnification and Artificial Neural Networks (ANN). The proposed approach creates a descriptor vector from processed Eulerian Magnification videos and uses an ANN classification to achieve a suitable accuracy rate for isolated and combined emotions. To validate the efficiency of the emotion classifier, confusion matrices were generated with the emotion classified by the emotion classifer and the real emotion of the video, tests were made with several sets of emotions. The classifier was abble to classify effitiently sets with 3 classes of emotions, confusing when more emotions were used

Keywords: Emotion detection, Face analysis, Eulerian Video Magnification.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Pesquisas usando detecção de emoções por meio da face envolvem fazer um computador ser capaz de identificar o estado emocional de um usuário humano. Esta área de pesquisa vem ganhando bastante atenção recentemente por conta do seu potencial de aplicação em diversos campos, tal como medicina [18], psicologia [41], robótica [35] e aplicações forenses [10], como um detector de mentiras. De acordo com [32], um comportamento facial dinâmico possui uma fonte rica de informações de transmissão de emoções. Como em qualquer comunicação, muita informação pode ser inferida da mensagem pela expressão facial de quem a emite. Estas características trazem à tona as tarefas de reconhecimento facial um desafio interessante para as indústrias e pesquisas de visão computacional. Muitos métodos foram usados por diferentes pesquisadores em seus esforços para alcançar uma abordagem eficiente que permita detectar emoções faciais humanas usando técnicas baseadas em visão computacional.

Suportado por argumentos anteriores, neste trabalho foi proposto o desenvolvimento inicial de um sistema robusto capaz de realizar uma classificação de emoções automática usando Magnificação de Vídeo Euleriana(MVE) e Redes Neurais Artificiais(RNA). A Seção 1.2 descreve os principais trabalhos relacionados com detecção automática de emoções. Nos Capítulos de Metodologia e Resultados, a metodologia proposta e os resultados iniciais são apresentados, respectivamente. Conclusões e futuros trabalhos são descritos no Capítulo Conclusão.

1.1 Objetivos e Justificativas

Nesta seção serão descritos os objetivos deste trabalho e as justificativas que levaram a realização deste, bem como estudos que foram feitos para iniciar a classificação e os métodos realizados para se testar e validar o classificador.

1.1.1 Objetivos Gerais

Este trabalho propõe uma abordagem para o desenvolvimento de um classificador automático de emoções. O classificador desenvolvido não possui nenhum conhecimento prévio do comportamento das pessoas em cada emoção e será treinado automaticamente a partir de vídeos com emoções conhecidas, gerando descritores para cada emoção do conjunto de emoções. O classificador deverá poder receber conjuntos de emoções de diferentes quantidades de classes de emoções, neste trabalho foram usados conjuntos de duas a seis emoções. Após o treinamento do classificador, este deve ser capaz de, a partir de um vídeo com a emoção desconhecida, identificar qual emoção a pessoa está expressando no vídeo.

1.1.2 Objetivos Específicos

Alguns objetivos específicos foram seguidos para construir o sistema de classificação de emoções. Entre eles, análise de artigos e técnicas de classificação foram feitos para gerar o sistema.

Estudar técnicas de Magnificação Euleriana foi o primeiro passo para gerar o sistema de reconhecimento de emoções. Para isto, os artigos referentes à MVE foram lidos e o código da MVE foi baixado e estudado e diversos fluxogramas para suas funções foram gerados para auxiliar no estudo. Para os estudos, os diversos tipos de magnificação foram testados, as variáveis geradas foram observadas, bem como as matrizes de filtros usadas para cada um dos tipos de magnificação. Foram feitos testes também para os diversos tipos de entrada com diferentes vídeos.

A segunda etapa foi estudar técnicas de reconhecimento de padrões de emoções em imagens e vídeos. Diversos artigos acerca do tema de reconhecimento de emoções, como Detecção de mmentiras usando micro-expressões faciais[23],Reconhecimento de emoções usando micro-expressões de corpo e face [40] e Análise de emoções usando bases multimodais [21], foram lidos, bem como artigos a acerca do reconhecimento de movimentos, como o uso de Histogramas Orientados ao Fluxo Óptico, *Histogram of Oriented Optical Flux*(HOOF) [31], a análise de movimento usada para classificar diferentes estilos de natação, visto em [39] e a análise de pontos específicos da face, usando a *framework* vista em CI2CV[3].

A partir dos estudos feitos em reconhecimento de emoções, diversos testes foram feitos para tentar gerar classificadores distintos que possam ser utilizados para a classificação de emoções. Cada uma das técnicas estudadas foi então testada em vídeos magnificados e os descritores gerados por estas técnicas foram analisados para se verificar se era possível gerar descritores para a classificação de emoção.

Para usar os descritores gerados anteriormente, diversas RNAs foram treinadas usando o vetor com os descritores como entrada. RNAs com diferentes quantidades de neurônios e camadas foram geradas e testadas. Primeiramente a rede foi treinada com vídeos com a emoção conhecida, e depois a rede foi usada para classificar vídeos onde a emoção era desconhecida e validado com a emoção real daquele vídeo para testar a eficiência da RNA naquele conjunto de emoções.

Após os testes simples, foram feitos testes com uma base de dados pública de emoções, realizando testes massivos para diferentes conjuntos de emoções da base. Matrizes de confusão foram geradas para verificar se o sistema consegue classificar eficientemente a emoção.

1.1.3 Justificativas

O processo de reconhecimento de emoções pode ser usado em diversas aplicações, sendo uma área de interesse para diversos setores, como Medicina, Vida cotidiana, Terapia, Educação, Monitoramento, Entretenimento e Marketing.

Em sistemas baseados de interface de cérebro e computador [28] é dito que o reconhecimento de emoções pode ser usado em medicina, é possível notar então que o reconhecimento de emoções pode auxiliar na reabilitição de um paciente, usando-o para monitorar o paciente. Pode ser usado também para o aconselhamento, em psicologia, podendo obter o estado emocional de um paciente através do reconhecimento de emoções. O reconhecimento de emoções pode ainda ser usado na assistência médica, obtendo os sentimentos do paciente em relação ao seu tratamento.

Em reconhecimento de emoçoes em tempo real[33], é apresentada uma aplicação de reconhecimento de emoções que pode ser usada na vida cotidiana, reconhecendo emoções em tempo real em pessoas que o usuário do dispositivo encontra na rua, isto pode usado para saber como lidar com pessoas de acordo com o que estas estão sentindo.

Estudos de reconhecimento de emoções [42] mostram que O reconhecimento de emoções pode ainda ser usado em terapias, reconhecendo o estado emocional do paciente e auxiliando este a lidar com stress, ansiedade e depressão.

Em sistemas inteligentes de tutoria[25], é mostrado que é possível ainda utilizar o reconhecimento de emoções no auxílio do aprendizado, podendo detectar o estado emocional do estudante e ajustar o estilo do exercício ou da apresentação de um tutor *online* conforme o estado emocional detectado, podendo também aumentar a interatividade no ensino a distância, provendo um *feedback* melhor para cada estudante de acordo com seu estado e aumentando assim a eficiência do aprendizado.

Reconhecimento de emoções a partir da fala, vistos em [8] e [42] mostram aplicações de reconhecimento de emoções sendo usadas em *call-centers* para detectar a emoção do

cliente e atendê-lo conforme o seu estado emocional, por exemplo, priorizando clientes que estão com raiva. É possível usar esse monitoramento também em caixas eletrônicos, impedindo pessoas assustadas de sacar dinheiro na tentativa de evitar assaltos. Segundo [13], é possível ainda usar o monitoramento em relacionamentos a distância para detectar o verdadeiro estado emocional do parceiro, como em aplicações para serviços móveis [43].

Como pode ser visto em [20], o classificador pode ser usado também para ter uma maior interação com o usuário. Em filmes ou jogos, é possível usá-lo para obter a emoção de quem está vendo um filme, jogando ou ouvindo música e modificar algo neste dependendo da emoção da pessoa, podendo por exemplo, mudar a música quando a pessoa fica nervosa,

A empresa *emotient* [9] mostra que é possível ainda usar o reconhecimento de emoções para medir o impacto de anúncios nas pessoas, podendo então produzir anúncios com maior poder para fazer com que as pessoas decidam comprar o produto, podendo então usar o classificador na área de *Marketing*.

1.2 Trabalhos Relacionados

A classificação de emoções está, presentemente, recebendo muita atenção pelos pesquisadores devido a seu potencial em diversas áreas de pesquisa como psicologia, estudos de comportamento humano e interface humano-computador. Atualmente, o foco principal está concentrado em sistemas que realizam a classificação baseada em informações da voz humana [1][30], combinados com características extraídas de uma face humana capturada por uma câmera, como descrito em [34].

O Sistema de Codificação de Ações Faciais, ou the *Facial Action Coding System* (FACS) [7] é o mais conhecido e o mais usado sistema desenvolvido por observadores humanos para descrever atividade facial em termos das ações musculares faciais visualmente observáveis (i.e., Ações de Unidades, *Action Units*, AUs). Com as FACS, observadores humanos unicamente decompõe expressões faciais em uma ou mais das 44 AUs que produziram a expressão em questão. De acordo com [27], muitos trabalhos tem aplicado diversas técnicas, como Redes Neurais Artificiais, Máquinas de Vetores de Suporte e Redes Bayesianas, para alcançar um sistema eficiente de classificação de emoção.

Em [2] é apresentada uma abordagem em tempo real para reconhecimento de emoções usando um buscador de características faciais automático para realizar a localização da face e extração de características. Os deslocamentos de características faciais no vídeo são usados como entrada para um classificador de Máquina de Vetor de Suporte. Eles avaliam o método proposto em termos da acurácia do reconhecimento para uma variedade de cenários de interação. Em [17], um método de segmentação que possa manusear imagens degradadas adquiridas em condições piores do que as propostas. Neste trabalho, eles

primeiramente usam informações dos elementos dos olhos (i.e. esclera), e um novo tipo de característica que mede a proporção entre os elementos do olho em cada direção que foi avaliado. Finalmente, após o processamento dos elementos do olho, o processo é viável para aplicações em tempo real.

Em outras técnicas, como na abordagem proposta descrita em [11], os autores focaram em detecção de emoções em falas raivosas-neutras, que são comuns em estudos recentes de Detecção de Variação de Emoção Automática, ou *Automatic Emotion Variation Detection* (AEVD). Este estudo propõe uma estrutura nova para AEVD usando Janelas Deslizaveis Multi-escaláveis, ou *Multi-scaled Sliding Window* (MSW- AEVD) para atribuir uma classe de emoção para cada mudança de janela por decisões de fusões de todas as janelas deslizantes contidas na mudança. Usando a estratégia de fusão paralela, eles alcançam uma efetividade de aproximadamente 92%. Infelizmente, nesta técnica proposta nenhuma informação adicional é descrita. Essa informação é crucial para explicar a alta taxa de eficiência obtida.

No estudo proposto por Byung-Hun e Kwang-Seok Hong[26], uma região facial é detectada através da combinação do sistema de cores $YCbCr$ com a imagem da Combinação do Gradiente Morfológico Máximo, ou *Maximum Morphological Gradient Combination* (MMGC). A região de pesquisa para a detecção do componente facial é limitada a região da face detectada. Os componentes faciais são detectados usando o método de histograma, o método de marcação de partículas e a imagem MMGC. Neste trabalho, os autores usaram um conjunto de dados de emoção especializado (eNTERFACE'05) para avaliar a metodologia proposta e foi alcançada uma acurácia de aproximadamente 81%.

Até agora, todos os trabalhos citados vistos na literatura científica resolveram (ou propuseram uma solução definitiva) o problema de reconhecimento de emoções. No entanto, entre os trabalhos existentes, uma abordagem inicial e viável está sendo proposta neste trabalho endereçada no aperfeiçoamento do problema de reconhecimento de emoções. Particularmente usando uma técnica onde todas as análises são feitas no rosto capturado de um vídeo e não requer o uso de nenhum outra informação, como a voz humana, para realizar o processamento do reconhecimento de emoções.

A partir das informações vistas anteriormente, no Capítulo 2 será feita uma revisão bibliográfica mostrando a teoria por trás das técnicas utilizadas para a realização deste projeto, mostrando como as técnicas funcionam. No Capítulo 3 é mostrada a metodologia do projeto, como o classificador funciona e todos os passos que o classificador faz para classificar a emoção. No Capítulo 4, são evidenciados os resultados do classificador, gerando matrizes de confusão para validar a eficiência do classificador e classificar as emoções. E no Capítulo 5 são discutidas as conclusões acerca da eficiência do classificador e são citados possíveis trabalhos futuros a serem realizados para melhorar o classificador.

Capítulo 2

Revisão Bibliográfica

Para a realização deste projeto, são necessários conhecimentos teóricos em outros projetos, como o de MVE e de técnicas de identificação de classificação de classes de vídeos. Estes conhecimentos foram usados como base para montar a ideia e para auxiliar a realização da implementação correta do projeto.

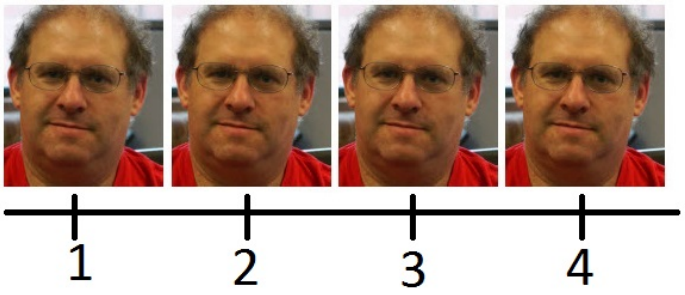
2.1 Magnificação de Vídeo Euleriana

A técnica que inspirou esse trabalho foi a MVE [38]. Esta técnica consiste na ampliação de movimentos do vídeo a partir de filtros que são passados em um vídeo de entrada, identificando e visualizando pequenas variações de *pixels* e usando-as para criar classificadores que não seriam possíveis com vídeos normais. Os vídeos magnificados amplificam não somente a variação de movimentos do vídeo, como também a de cor, sendo possível, por exemplo, verificar as pulsações de uma pessoa provocadas pela corrente sanguínea. Como exemplo de magnificação baseada em cor, temos a Figura 2.1, que mostra os frames do vídeo ao longo do tempo.

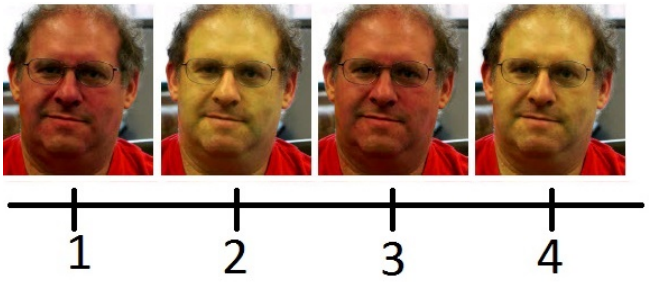
A Figura 2.1 mostra 4 *frames* do vídeo de entrada em (a) e os mesmo 4 *frames* do vídeo de saída em (b). Após a magnificação, é perceptível a diferença de cor no rosto da pessoa. Na Figura 2.2, é notável a visualização das fatias espaço-temporais de variação de cor dos vídeos de entrada e saída do vídeo visto na Figura 2.1. O gráfico (a) é o gráfico de entrada, pego no vídeo original, e o gráfico (b) mostra as fatias espaço-temporais magnificadas. Nas fatias do vídeo de entrada a variação de cor é imperceptível, enquanto que no vídeo de saída, as variações de cor são facilmente perceptíveis.

Além da magnificação de cor, temos também a magnificação de movimento. Um exemplo de movimento é apresentado na Figura 2.3.

Na Figura 2.3 vemos um guindaste balançando no vento. Na imagem da esquerda, temos o *frame* completo extraído do vídeo. Nas imagens de cima temos um corte da



(a) Vídeo original



(b) Vídeo magnificado

Figura 2.1: Exemplo de um vídeo ao lado do mesmo vídeo magnificado. Adaptado de [38].

imagem do guindaste magnificado e nas imagens de baixo temos as fatias espaço-temporais do corte do guindaste.

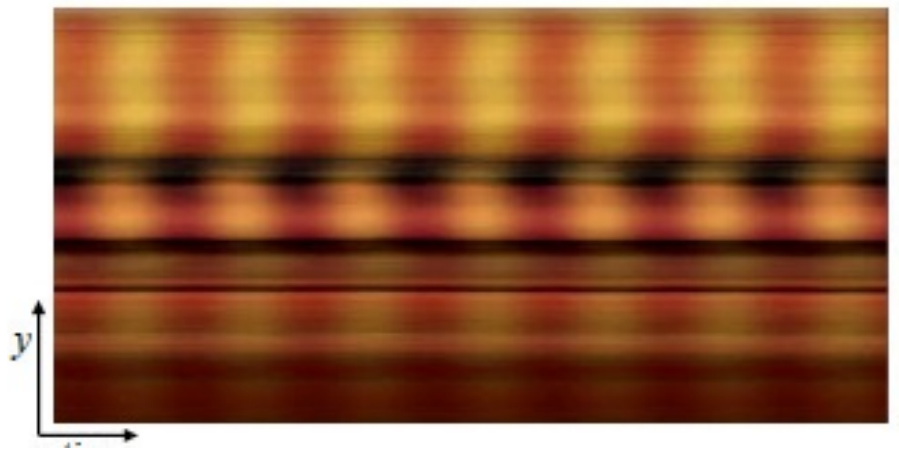
2.1.1 Processamento

Segundo [38], a abordagem básica consiste em considerar os valores de cor em um período de tempo em uma determinada localização espacial e amplificar a variação desses valores em uma dada faixa de frequência de interesse. Para a amplificação vista na Figura 2.2, por exemplo, o filtro foi aplicado para baixas frequências.

A filtragem temporal não somente amplifica variações de cor, mas consegue também amplificar pequenas variações de movimento. Como por exemplo, em [38] é possível ver a magnificação sendo usada para amplificar o peito de um bebê respirando.

A técnica de MVE é baseada em um processamento espacial-temporal. Na Figura 2.4 podemos ver um fluxograma do processamento geral. Primeiramente, o sistema decompõe o vídeo de entrada em diferentes sequências de faixa de frequência espaciais e aplica o mesmo filtro temporal em todas as faixas. As faixas espaciais filtradas são então amplificadas por um fator *alpha* (α), que é adicionado ao sinal original, usado para gerar o vídeo de saída. A escolha de filtro temporal e dos fatores de amplificação podem ser alteradas para diferentes aplicações.

(a) *Slices* do vídeo de entrada



(b) *Slices* do vídeo magnificado

Figura 2.2: Exemplo dos *slices* de um vídeo de entrada ao lado dos *slices* do mesmo vídeo magnificado. Adaptado de [38].

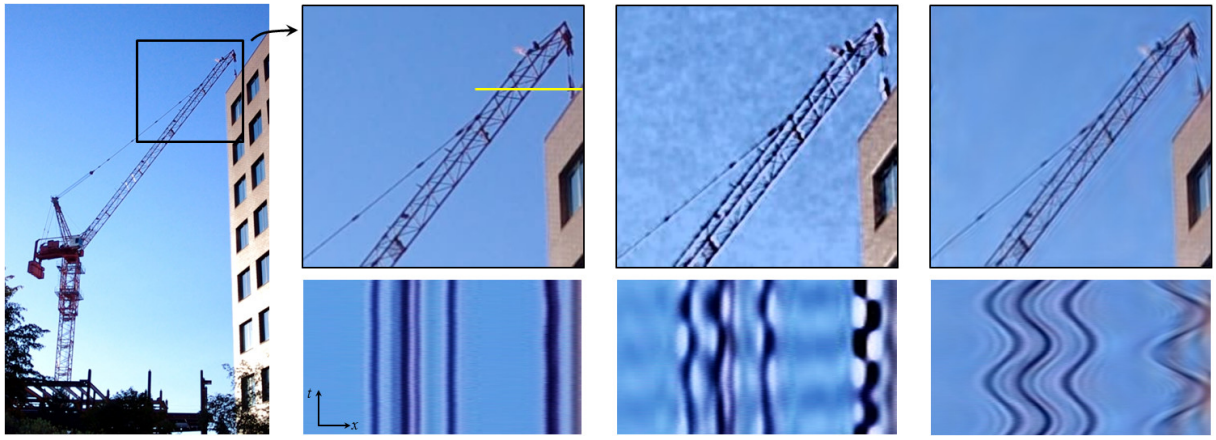


Figura 2.3: Exemplo de magnificação de movimento, retirado de [37].

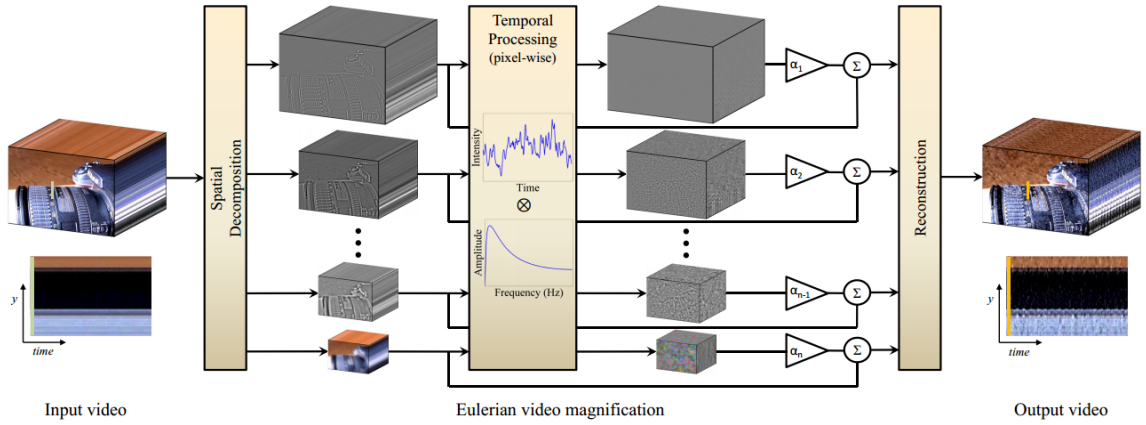


Figura 2.4: Fluxograma da magnificação retirado de [38].

2.1.2 Processamento Espacial

Em [38], é possível ver que a técnica possui opções de processamento espacial e temporal. Para as opções de processamento espacial temos a possibilidade de usar técnicas Laplacianas e Gaussianas. O processamento com técnicas Laplacianas pode ser realizado com diferentes tipos de processamentos temporais, enquanto o processamento Gaussiano permite somente o processamento temporal ideal.

O motivo pelo qual o processamento Gaussiano pode ser realizado somente com o processamento temporal ideal é que este não consegue ser realizado *frame* a *frame* a medida que o vídeo é lido, não podendo ser realizado com pirâmides, sendo necessário criar uma pilha de filtros gaussianos para o vídeo antes de processá-lo, ou seja, o vídeo é percorrido inteiramente, criando pilhas de filtros, e então é percorrido novamente *frame* a *frame* usando os filtros obtidos na pilha Gaussiana.

A técnica de processamento espacial Laplaciana permite, além do processamento temporal ideal, outros tipo de processamentos temporais, como o *Butter* e o *Infinite Impulse Response* (IIR), que usam respectivamente os filtros *ButterWorth* e *Infinite Impulse Response* (IIR), e que criam pirâmides laplacianas enquanto percorrem o vídeo e passam os *frames* no filtros, ou seja, os vídeos são percorridos somente uma vez.

Já a técnica de processamento espacial Gaussiana utiliza, para criar os filtros que serão utilizados no vídeo original, uma técnica de borragem e reamostragem da imagem original, usando um núcleo chamado *binom5*, visto em 2.2, o filtro então é criado realizando diversas correlações entre a imagem de entrada e o núcleo.

A correlação entre as imagens é feita convoluindo a matriz da imagem de entrada, o *frame* do vídeo original, e a matriz do núcleo escolhido, e então a matriz resultado é

reamostrada para um tamanho menor a partir dos *pixels* iniciais e finais em x e y, que recebe como parâmetro, e de um passo, que indica quantos *pixels* o laço vai percorrer em cada iteração para gerar a matriz final, ou seja, se o valor do passo em x for 2, a matriz final terá metade da largura da matriz antes da reamostragem. A correlação entre duas funções x e y pode ser vista na Equação 2.1, onde X e Y são as variáveis aleatórias, μ_x e μ_y as médias de x e y respectivamente, σ_x e σ_y os desvios padrões de e y respectivamente e E é o valor esperado.

$$coor(X, Y) = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y}. \tag{2.1}$$

$$Binom5 = [0.0884, 0.3536, 0.5303, 0.3536, 0.0884] \tag{2.2}$$

Para o processamento espacial Laplaciano, o algoritmo constrói pirâmides Laplacianas, e para isso primeiramente obtém o tamanho máximo da pirâmide a partir da Equação 2.3, onde *Maxht* é o tamanho máximo da pirâmide e *Size* é o tamanho da maior dimensão da imagem, largura ou altura.

$$Maxht = \log_2 Size \tag{2.3}$$

A partir do tamanho da pirâmide, o algoritmo faz uma série de reduções até chegar na base da pirâmide, onde então volta na pirâmide fazendo expansões nos filtros criados com a redução. Esses filtros expandidos são então guardados na pirâmide para serem usados para magnificar.

As reduções e expansões são feitas a partir de reamostragens dos filtros, reamostrando para um tamanho maior ou menor, sendo que quando reamostrada para um tamanho maior, os *pixels* são preenchidos com zeros.

2.1.3 Processamento Temporal

Ainda em [38], é visto que a MVE possui três tipos de processamento temporal. O processamento temporal ideal é aquele que percorre todo o vídeo gerando uma pilha de filtros antes de efetivamente percorrer o vídeo aplicando os filtros nos *frames*. Este método é compatível com o processamento espacial Laplaciano e o Gaussiano. No processamento temporal ideal, após a pilha de filtros ser criada é feita uma filtragem passa banda nessa pilha, extraindo somente a largura de banda desejada, e então o vídeo é repercorrido, sendo o filtro referente a cada vídeo extraído da pilha e somado no *frame* do vídeo original para gerar o vídeo de saída. Um fluxograma do processamento temporal ideal pode ser visto na Figura 2.5 .

Os outros dois tipos de processamento temporal são chamados pelo artigo de *Butter* e IIR, que são compatíveis somente com o processamento espacial Laplaciano, visto que é necessário que sejam criadas pirâmides Laplacianas para o processamento, onde o vídeo é percorrido somente uma vez, criando as pirâmides e filtrando os *frames*. A diferença entre o processamento temporal IIR e *Butter* está no modo como os filtros são criados e aplicados nos *frames*, onde o *Butter* usa a pirâmide do *frame* anterior para criar o filtro que será somado no *frame*.

2.2 Redes Neurais

Segundo *Neural Networks: A Comprehensive Foundation* [16] e *Neural Network Toolbox^{TM} User's Guide* [6], redes neurais são modelos estatísticos inspirados nas redes neurais biológicas, como o cérebro, usados para prever um resultado, uma saída, a partir de um número, normalmente alto, de valores de entrada após ser treinada com uma série de valores de entrada e saída esperados. Redes neurais podem ser representadas por um sistema com diversos neurônios interligados que trocam mensagens entre si, sendo que as conexões entre os neurônios possuem pesos numéricos, obtidos no treinamento da rede. Quando a rede treinada recebe uma entrada, esta passa por seus diversos neurônios até gerar sua saída.

Existem ainda as redes de multicamadas, que funcionam como se fossem várias redes independentes, onde a saída de uma é a entrada de outra.

Para se criar uma rede neural devemos primeiro coletar os dados necessários para criá-la, ou seja, valores de entrada e a saída esperada para estes. A rede deve ser então criada e configurada, escolhendo os valores como quantidade de parâmetros de entrada e saída, quantidade de camadas e de neurônios. É preciso então inicializar os pesos dos neurônios e treinar a rede com os dados coletados e suas saídas esperadas. Após a rede ser treinada, é passada por um processo de validação e está pronta para uso.

Redes neurais são compostas por diversos neurônios, cada neurônio consiste de vetores de entrada, que são multiplicada por pesos e depois somada por um valor de *bias*, sendo então passado por uma função de transferência, e gerando uma saída, como podemos ver na Figura 2.6.

2.2.1 Redes estáticas x dinâmicas

Uma rede neural pode ser classificada como estática ou dinâmica. A saída de uma rede neural estática depende somente da sua entrada atual e não possui atrasos, ou seja, não depende de entradas passadas. Já uma rede dinâmica depende não só da entrada atual da rede, mas de entradas passadas, saídas passadas e estados passados da rede.

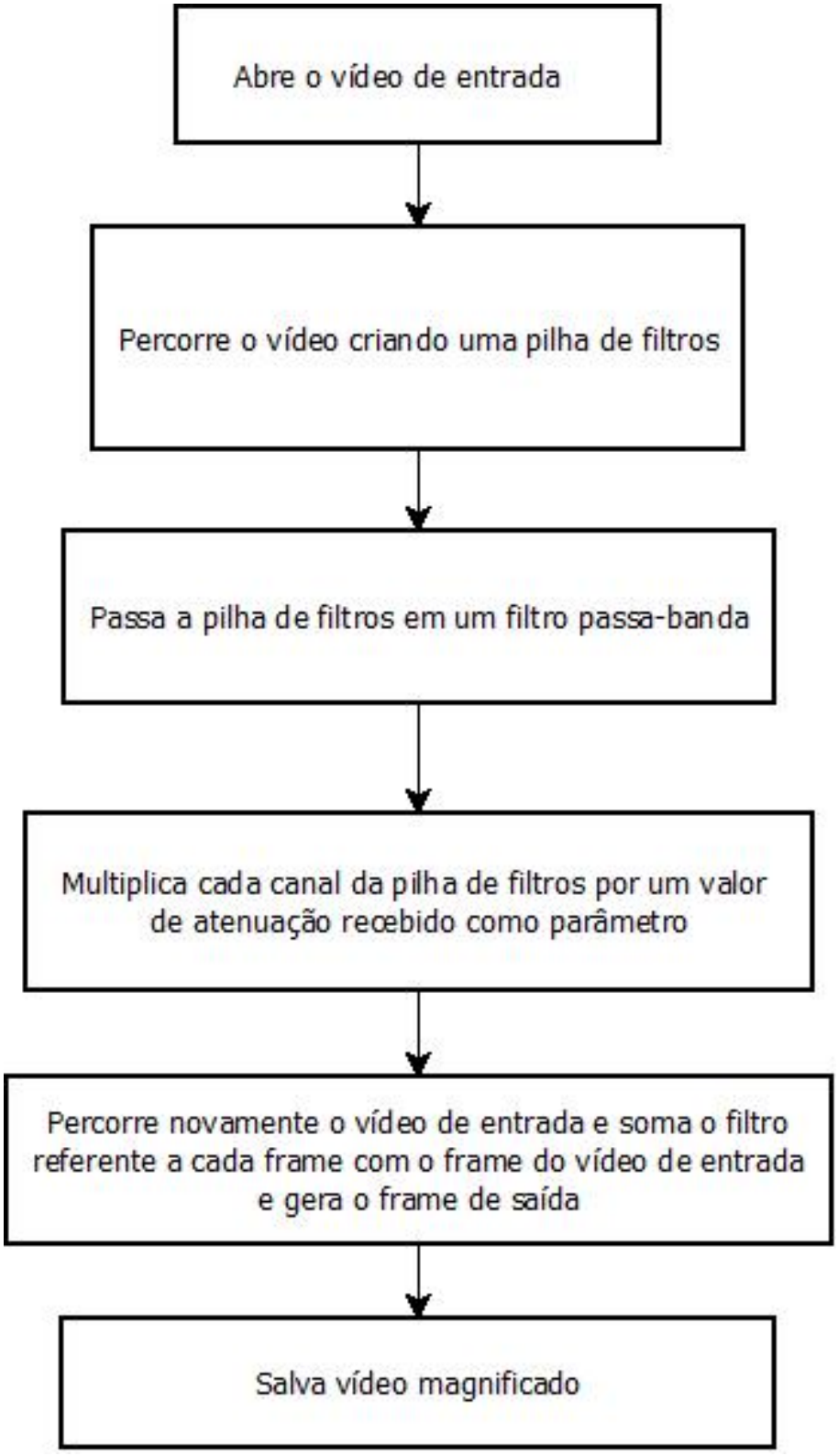


Figura 2.5: Fluxograma do processamento temporal ideal, feito no *Dia Diagram* [14] com as informações obtidas a partir do código no artigo MVE[38]..

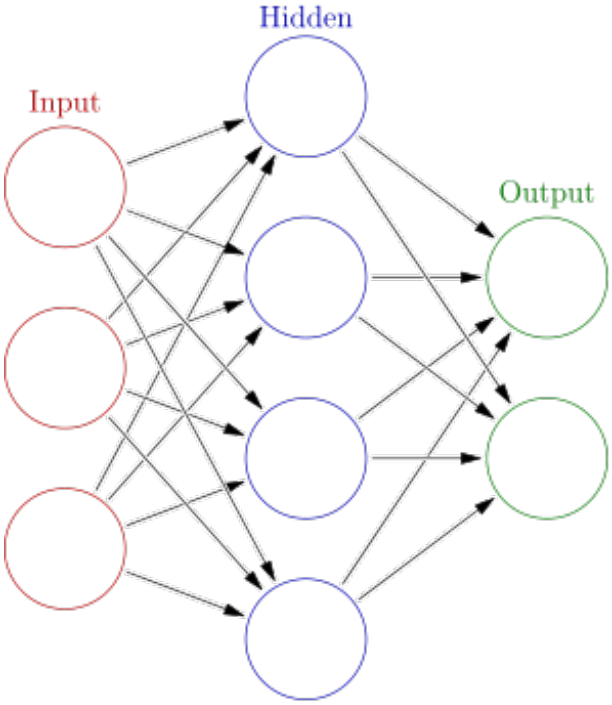


Figura 2.6: Exemplo de uma RNA, extraída de [15].

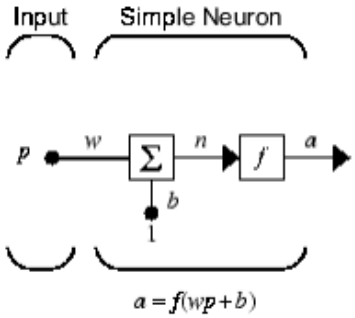


Figura 2.7: Exemplo de um neurônio simples, extraído de [6].

2.2.2 Treinamento *batch* x *adapt*

As redes neurais podem ser treinadas de duas formas, temos o treinamento *batch* e o treinamento *adapt*. No treinamento *batch* a rede é treinada previamente com diversos valores de entrada e saídas esperadas, ou seja, a rede vai receber diversos valores de entrada e saída de uma vez só e treinar. Esta é a forma mais rápida de treinamento, porém, pode ser feita somente uma vez, antes de se usar a rede. Já no treinamento *adapt*, é possível treinar a rede com novos parâmetros diversas vezes, ou seja, podendo ainda retreinar a rede para novos parâmetros sempre que desejado. Este método é mais lento

que o *batch*, porém, não há necessidade de criar novas redes com o passar do tempo, simplesmente retreinar.

2.3 Descritores estatísticos

Os descritores estatísticos utilizados foram média, variância, obliquidade e curtose. Descritores estatísticos são usados em estatística para obter valores de funções de somente uma variável, sendo unidimensionais. Porém, uma imagem é bidimensional, uma imagem de altura M e largura N terá MxN *pixels*, e portanto MxN valores para serem usados nas funções estatísticas, dado que o valor do canal de cor daquele *pixel* é o usado na função, nesse caso o canal Luma Y do sistema de cores $YCbCr$.

As funções dos descritores recebem um vetor de entrada para gerar os descritores, para utilizar essas funções com uma matriz ao invés de um vetor, a função é usada mais de uma vez. Primeiramente todas as colunas da matriz MxN da imagem são percorridas, obtendo N vetores de tamanho M, e para cada um desses vetores é obtida um valor de descritor, gerando então N descritores para cada imagem. Com esses N descritores, um vetor de tamanho N é montado e o mesmo descritor é aplicado nesse vetor, obtendo finalmente um descritor para cada imagem, ou cada *frame*.

Como cada vídeo possui um descritor e não cada *frame*, o vídeo é percorrido e um descritor é gerado para cada *frame*, depois a média dos descritores é tirada e o descritor do vídeo é obtido.

2.3.1 Média

O descritor estatístico média (μ) pode ser visto na Equação 2.4, é de fato a média de todas as variáveis X_i da função estatística.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{2.4}$$

A média aponta para o valor esperado para o qual os valores das variáveis da função se concentram, sendo então o valor significativo da função. Neste projeto, a média fornece o valor significativo da componente Luma para o vídeo.

2.3.2 Variância

A variância mede o quanto as variáveis X_i de uma função estatística estão espalhadas, quando o valor da variância é zero significa que todos os valores são idênticos. É calculada usando a Equação 2.5.

$$Var = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}. \tag{2.5}$$

A variância neste trabalho irá mostrar o quanto a componente Luma Y variou no vídeo sendo processado, medindo a dispersão da variação da grandeza Luma(Y). A variância irá indicar o quanto a pele irá variar no vídeo, cuja variação será amplificada com a magnificação. Portanto, uma variância de valor elevado significa que a cor da pele da pessoa variou muito, mostrando que a reação da pessoa a emoção foi forte ou perceptível.

2.3.3 Obliquidade

Obliquidade é a medida de simetria, ou assimetria, de uma distribuição probabilística, O valor da obliquidade pode ser positivo, negativo, ou até mesmo indefinido, sendo que uma distribuição simétrica tem obliquidade de valor zero. Se uma distribuição tem obliquidade positiva, então tem uma longa calda para a direita, em seus valores mais altos, e se tem valor negativo, então tem uma longa calda para esquerda, em seus valores mais baixos. A obliquidade pode ser calculada segundo a Equação 2.6 a seguir:

$$O = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^3}{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2)^{3/2}}. \tag{2.6}$$

Onde O é o valor da obliquidade e μ é a média, σ é o desvio padrão, X é a variável aleatória da função da qual a obliquidade está sendo obtida e E é o valor esperado.

2.3.4 Curtose

Curtose mede o quanto uma distribuição se assemelha com a distribuição Gaussiana, a partir dos picos dessa distribuição. Uma distribuição com valor de curtose positiva tem mais picos que a distribuição gaussiana e uma com valor de curtose negativa é mais lisa que a distribuição gaussiana. A curtose é calculada usando a Equação 2.7 abaixo

$$C = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2)^2} - 3. \tag{2.7}$$

Onde C é o valor da curtose, x_i é a variável aleatória da função da qual a curtose está sendo obtida e μ é a média.

2.4 Cores

Segundo [29], a utilização de cores no processamento de imagens é motivada por diversos fatores, entre eles podemos citar o fato da cor ser um poderoso descritor, podendo ser usada na identificação e extração de objetos. Outro motivo é o fato de os humanos conseguirem discernir milhares de tons e intensidades de cores enquanto conseguem diferenciar aproximadamente doze tons de cinza. Estes fatores por si só justificam a utilização de cores na análise de imagens.

Algumas técnicas usadas em imagens em escalas de cinza podem ser diretamente aplicadas em imagens com cor, enquanto outras precisam de uma reformulação para serem consistentes com as propriedades do espaço de cores.

2.4.1 Fundamentos de cores

O processo do cérebro humano de perceber e interpretar cores, assim como o de outros animais, é um fenômeno psicológico ainda não entendido por completo, porém, a natureza física das cores pode ser expressada por bases formais por meio de resultados teóricos e experimentais.

Na natureza, as cores são formadas pela variação de frequência da luz branca. Quando a luz solar branca passa por um prisma, temos do outro lado os raios de cores formados pela saída da luz, indo do violeta ao vermelho, ou seja, a luz branca é decomposta em diversas outras cores. Podemos dividir o espectro de cores de saída em sete regiões: violeta, azul, anil, verde, amarelo, laranja e vermelho. A cor de um objeto é determinada pela resposta perceptiva do ser humano ao espectro de frequência da radiação incidente no olho

resposta perceptiva do ser humano ao espectro de frequência da radiação incidente no olho

Na Figura 2.8 podemos ver o experimento realizado por Isaac Newton em 1672, onde a luz branca atravessou o prisma e foi decomposta em sete feixes de cores diferentes.

Nos humanos, os sensores responsáveis pela visão das cores são os cones. Evidências experimentais mostram que o olho humano tem de 6 a 7 milhões de cones, e estes podem ser divididos em três categorias principais: as cores verde, vermelho e azul. Portanto, as cores vistas pelos humanos são combinações dessas três cores, de tal forma que a presença dessas três cores vai resultar na cor branca, e q ausência das três vai resultar na cor preta.

Figura 2.8: Luz branca sendo decomposta em outras cores, retirada de [4].

2.4.2 Sistemas de cores

As imagens digitais podem ser reproduzidas com diferentes tipos de sistemas de cores, que possuem diferentes canais de cores. O sistema de cor mais comum é o RGB, que é similar ao modo como os olhos humanos percebem as cores e é a abreviatura para Vermelho(*Red*), Verde(*Green*) e Azul(*Blue*). Também chamado de sistema aditivo, este sistema de cores monta a imagem somando os 3 canais de cores, o valor zero em todos os canais representa a cor preta e o valor 255, maior valor, em todos os canais, representa a cor branca.

Outro sistema de cores bastante utilizado, principalmente por impressoras, é o CMYK, que é composto pelos canais de Ciano(*Cian*), Magenta(*Magenta*) e Amarelo(*Yellow*), este sistema de cores é chamado também de subtrativo por ser contrário ao sistema RGB, ou seja, quando todos os canais tem valor zero o sistema representa a cor branca e quando todos os canais tem valor 255 o sistema representa a cor preta.

Um sistema de cores também bastante utilizado é o sistema Y C_b C_r, onde Y representa o valor de Luma, que representa o brilho de uma imagem, C_b representa a diferenciação de azul e C_r representa a diferenciação de vermelho.

2.4.3 Espaço de Cores YCbCr

O espaço de cores Y C_b C_r, é bastante usado na em imagens e vídeos digitais e não é um espaço de cores absoluto, é um modo de codificar informação RGB e as cores reais dependem do RGB usado para gerar o sinal. Portanto, um valor de Y C_b C_r só pode ser previsto se seus valores primários de RGB forem usados.

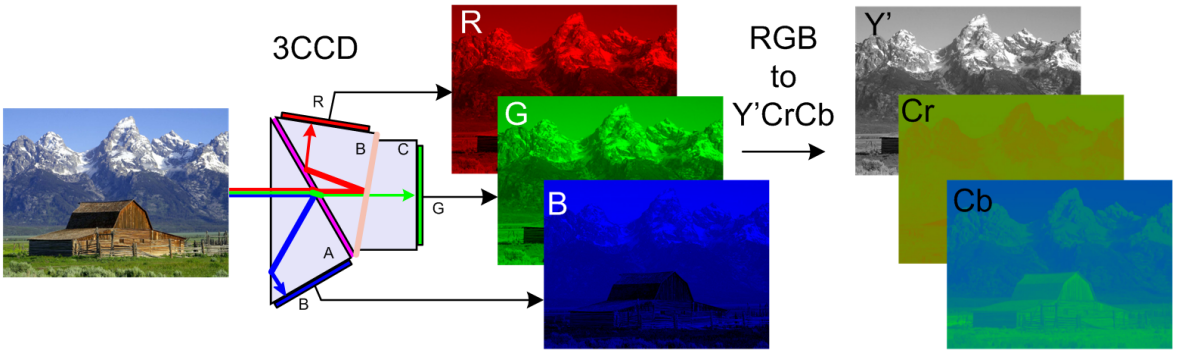


Figura 2.9: Exemplo da conversão de uma imagem de RGB para Y C_b C_r, retirada de [19] .

Na Figura 2.9 pode-se ver um fluxograma exemplificando a extração do sistema de cores RGB de uma imagem e sua conversão para o sistema de cores Y C_b C_r e na Figura 2.10, a imagem original juntamente com seus canais de cores Y C_b C_r gerados.

Segundo [12], o espaço de cores Y C_b C_r é bastante influenciado pela pele, portanto, como a MVE trabalha nas variações de movimento e cor, esse será bastante influenciado pelo movimento da pele da pessoa que está sendo analisada, obtendo-se uma maior variação perceptível no espaço de cores Y C_b C_r.

Neste cápitulo foram mostradas e explicadas as técnicas e teorias que serão usadas neste trabalho, no cápitulo seguinte será mostrado a metodologia acerca do trabalho, mostrando como cada uma das técnicas apresentadas se insere nesta.

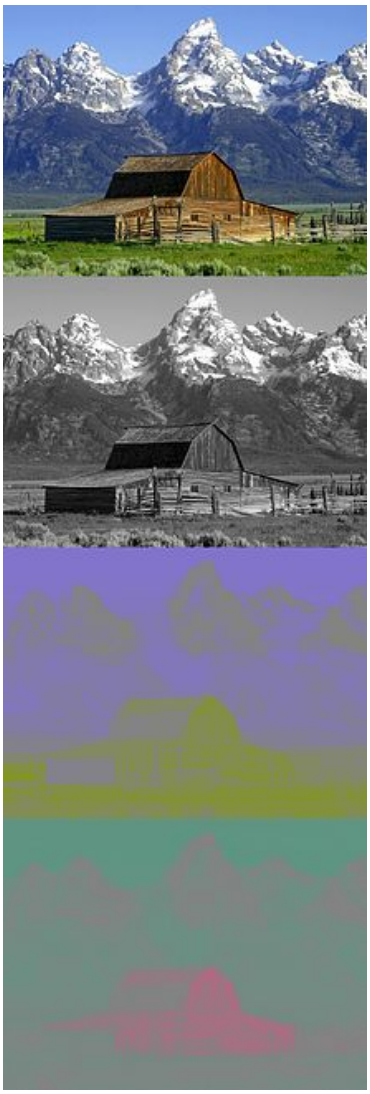


Figura 2.10: Exemplo de uma imagem decomposta no sistema de cores Y C_b C_r, retirada de [24] .

Capítulo 3

Metodologia

Durante a realização deste projeto diversas técnicas foram testadas e diversos algoritmos foram desenvolvidos. As técnicas e algoritmos usados neste trabalho serão apresentados nesta seção, bem como as bases de dados usadas.

Para se conseguir uma classificação de emoções automática, foi proposta uma metodologia baseada em MVE, descrita em [38]. Com essa técnica, é possível obter pequenas variações de cor e movimento em um vídeo. A amplificação traz mais informações na pessoa gravada, definidas como micro-expressões faciais, e estas podem ser usadas para gerar descritores e treinar RNAs que possuem a capacidade de classificar e reconhecer essas emoções. Os passos para se realizar a metodologia proposta estão descritos na Figura 3.1.

Na metodologia proposta, a única informação usada foi obtida pelo vídeo de uma pessoa que expressa uma emoção definida.

Para extrair a emoção, uma pessoa fala frases definidas que irão fazer com que certa emoção ocorra. Mesmo com o áudio disponível, a metodologia proposta não usa informação do áudio para analisar o vídeo.

Após abrir o vídeo, é feito um processo de detecção de face. Subsequentemente, o vídeo é magnificado usando o método de MVE, como em [38]. Após a magnificação, o vídeo é lido e é extraída a componente Luma Y do vídeo, convertendo cada *frame* do vídeo para o sistema de cores $\mathrm{Y}C_bC_r$. Com a componente Luma Y, foram gerados descritores que serão usados para treinar a rede neural que irá reconhecer as emoções.

Além de tratar dos métodos que efetivamente conseguiram reconhecer as emoções, trata-se também dos métodos que não obtiveram sucesso.

3.1 Tratamento dos dados de entrada

Antes de processar os vídeos de entrada, os vídeos passam por tratamentos para eliminar certas informações indesejadas, como o corte de face, que retira as informações de fundo

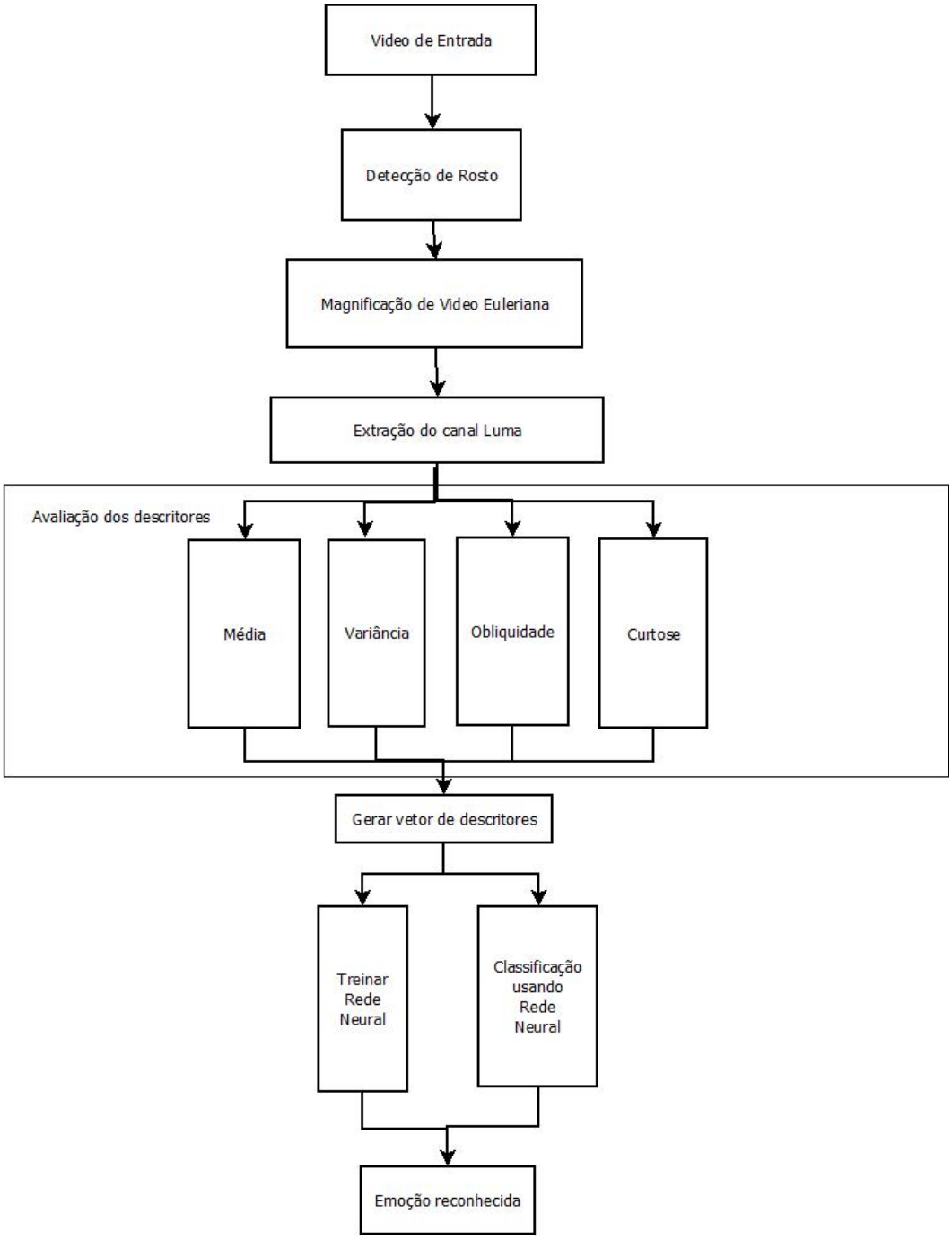


Figura 3.1: Fluxograma da metodologia feito no *Dia Diagram* [14].

e do corpo do sujeito em análise.

3.1.1 Corte de rosto

O corte do rosto consiste em um tratamento de entrada utilizado para se obter a região de interesse que será utilizada para processamento e retirar a região que não interessa para

o processamento. A região de interesse para processamento é o rosto da pessoa.

Utilizou-se um algoritmo de detecção facial, cortando a área detectada pelo algoritmo. Este processo resulta na geração de um novo vídeo somente com a área de interesse. O algoritmo usado para a detecção do rosto foi Viola-Jones com Haar-Cascade, descrito em [36].

Na Figura 3.2 é possível ver exemplos de um vídeo original da base ao lado do mesmo vídeo com o rosto cortado e na Figura 3.3 é possível ver um vídeo magnificado ao lado do mesmo vídeo com o rosto cortado. O vídeo usado para gerar os descritores é o vídeo magnificado com o rosto cortado.

(a) Vídeo original (b) Vídeo cortado

Figura 3.2: Exemplo de um vídeo original ao lado do mesmo vídeo com o rosto cortado

(a) Vídeo original (b) Vídeo cortado

Figura 3.3: Exemplo de um vídeo magnificado ao lado do meso vídeo com o rosto cortado

3.1.2 Magnificação de vídeo euleriana

A MVE [38] é a base para esse projeto, portanto, os vídeos usados de entrada passaram pelo processo de magnificação antes de serem usados como entrada para os métodos de identificação de emoções. A MVE é feita usando todas as informações do vídeo completo original e o vídeo magnificado é cortado usando os limites do vídeo com o rosto cortado anteriormente, dessa forma, nenhuma informação é perdida na magnificação.

O vídeo magnificado não pode ser usado diretamente do algoritmo de corte de rosto pois este não conseguirá identificar o rosto devido à alta taxa de ruído do vídeo magnificado.

O tipo de magnificação que teve melhores resultados e que será usado no sistema foi o com processamento temporal Ideal e processamento espacial Laplaciano.

(a) Vídeo original (b) Vídeo magnificado

Figura 3.4: Exemplo de um vídeo original ao lado de um vídeo magnificado

3.2 Processamento dos vídeos para classificação

Diversas técnicas foram usadas para processar os vídeos de entrada e tentar gerar classificadores para diferenciar os tipos de emoção. Muitos desses processamentos não geraram resultados que possibilitassem a diferenciação dos vídeos em classes de emoção.

3.2.1 Componente Luma

Após a MVE, o vídeo magnificado é usado como entrada, e o algoritmo percorre cada um dos seus *frames*, convertendo estes do sistema de cores RGB para o sistema de cores $YCbCr$. Após esse processo tem-se um valor de cada componente $YCbCr$ para cada *pixel* de cada *frame*. O componente Luma Y será usado para gerar descritores para a classificação das emoções.

Os componentes Cb e Cr tem valores muito similares para emoções distintas, ou seja, não são bons para serem usados como descritores para as emoções,. Portanto, esses componentes foram descartados como candidatos para os descritores no estágio de classificação das emoções.

3.2.2 Avaliação dos descritores

Com os valores de Luma para todos os pixels de cada *frame*, foram definidos quatro elementos para serem usados no vetor de descritores: média, variância, curtose e obliquidade. Quando esses valores são usados para analisar o vídeo magnificado, quatro variáveis para o estágio de classificação são obtidas.

As quatro variáveis foram obtidas usando as fórmulas apresentadas na Seção Revisão Bibliográfica. Dessa forma, cada vídeo terá um valor para cada uma dessas variáveis.

3.2.3 Vetor de descritores

Para o treinamento da RNA, foi gerado um vetor de descritores a partir das variáveis obtidas pelos descritores. O vetor criado contém os quatro valores, média, variância, obliquidade e curtose. Todas essas variáveis são normalizadas entre $[0, 1]$.

$$Vetor = [Média, Variância, Obliquidade, Curtose] \tag{3.1}$$

Na Tabela 3.1 é possível verificar que para uma pessoa cada emoção é distinta, quando usado o vetor de descritores proposto (Equação 3.1). Portanto, é possível utilizar estes descritores para classificar as emoções. Na Equação 3.2, é possível ver um exemplo do vetor de descritores criado usando valores da tabela.

$$Vetor = [0.6671, 0.1285, 1, 0.8591] \tag{3.2}$$

Valores	Média	Variância	Obliquidade	Curtose
Raiva	0.6671	0.1285	1	0.8591
Desgosto	0	0.1344	0.0188	0.0346
Medo	0.9709	0.0394	0.2032	0.6035
Felicidade	0.2937	0.4583	0.1962	0.6066
Tristeza	0.9797	0.6001	0.2928	0.8056
Surpresa	0.8718	0.1309	0.2362	0.7925

Tabela 3.1: Valores de média, variância, obliquidade e curtose de uma pessoa normalizados de 0 a 1 a usando todos os valores.

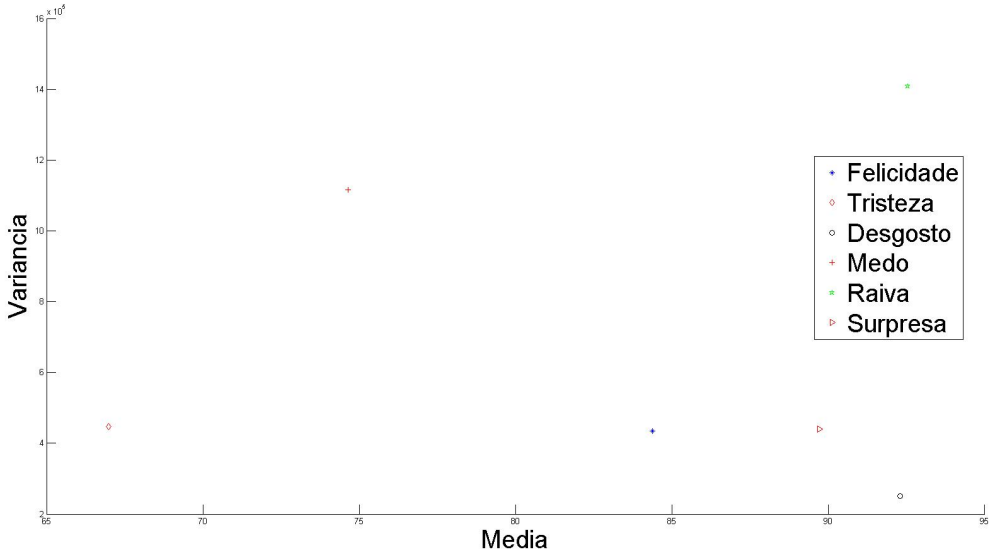


Figura 3.5: Descritores variância e média mostrados graficamente para seis emoções de um mesmo sujeito.

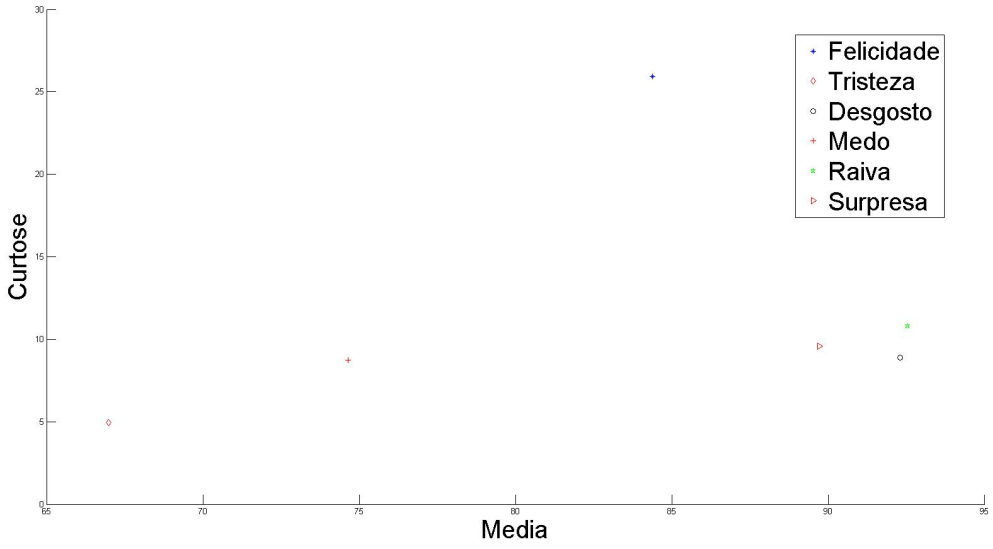


Figura 3.6: Descritores curtose e média mostrados graficamente para seis emoções de um mesmo sujeito.

Nas Figuras 3.5 a 3.10 foram mostrados graficamente todos os descritores para as seis diferentes classes de emoções de um mesmo sujeito. Analisando estes visualmente, é possível notar que o os descritores geram nuvens distintas, o que mostra que estes descritores podem ser usados para treinar uma RNA para classificar as emoções.

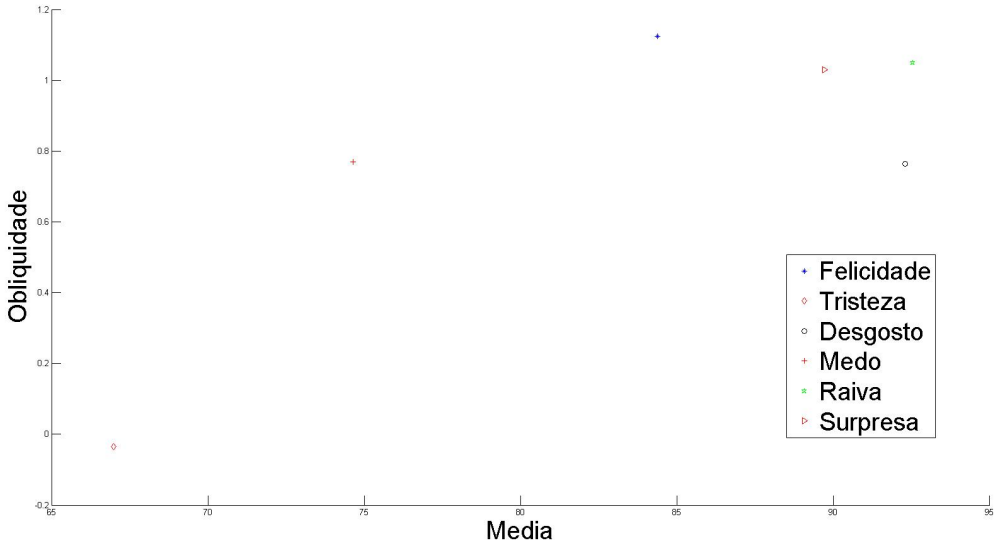


Figura 3.7: Descritores obliquidade e média mostrados graficamente para seis emoções de um mesmo sujeito.

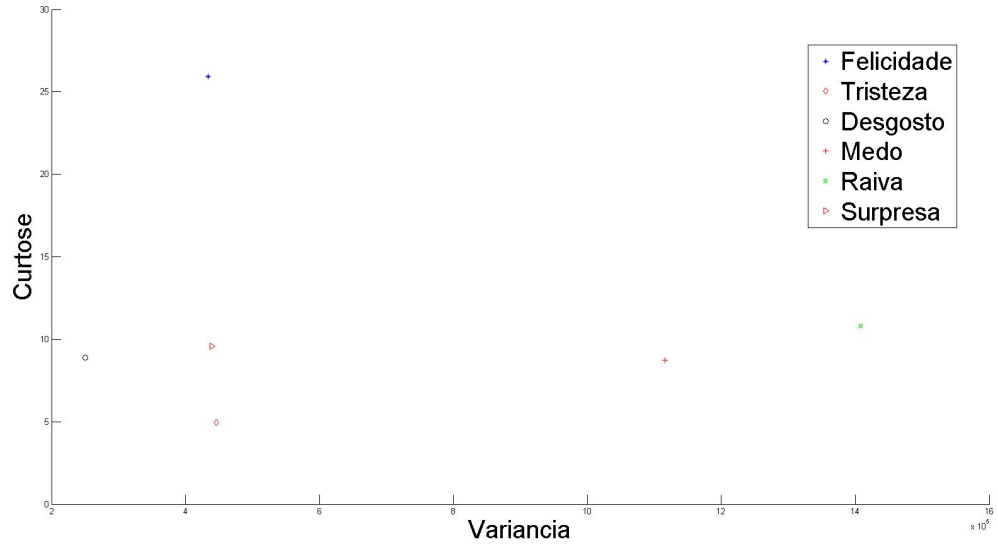


Figura 3.8: Descritores curtose e variância mostrados graficamente para seis emoções de um mesmo sujeito.

Porém, ao analisar as Figuras 3.11 a 3.16 é perceptível visualmente que com diversos sujeitos, as nuvens não ficam tão distintas, sendo então visto que para sujeitos diferentes, o classificador não irá conseguir usar estes descritores para classificar de forma eficiente.

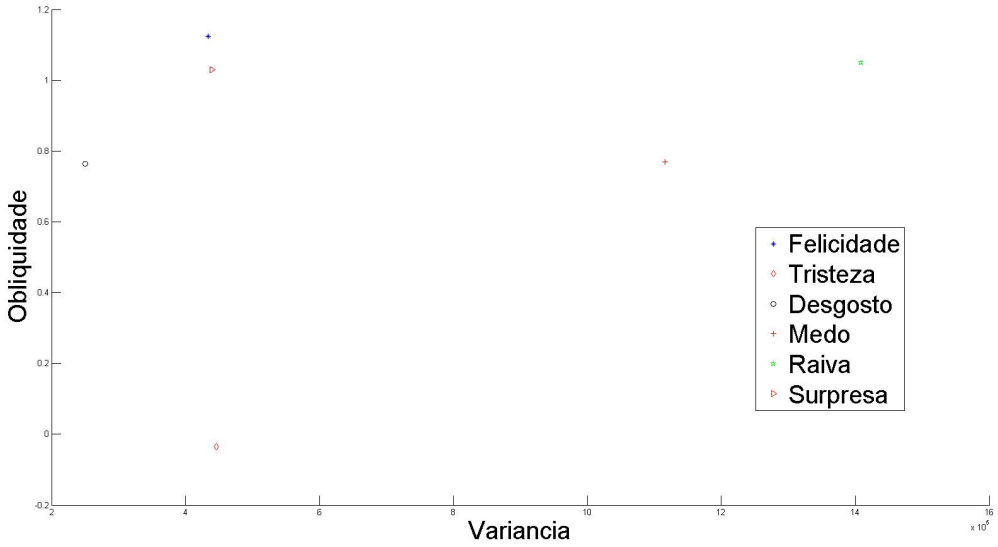


Figura 3.9: Descritores obliquidade e variância mostrados graficamente para seis emoções de um mesmo sujeito.

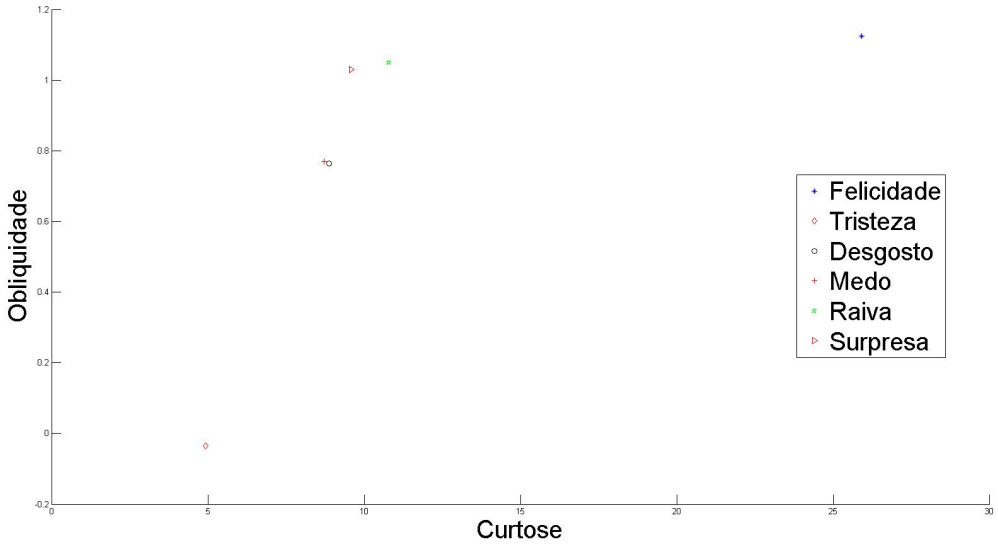


Figura 3.10: Descritores obliquidade e curtose mostrados graficamente para seis emoções de um mesmo sujeito.

3.2.4 Rede Neural Artificial

Para utilizar o vetor de descritores proposto para classificar as emoções, a RNA foi treinada com um conjunto reduzido de vídeos dos quais tem-se a informação prévia de qual é a emoção expressada nesses vídeos.

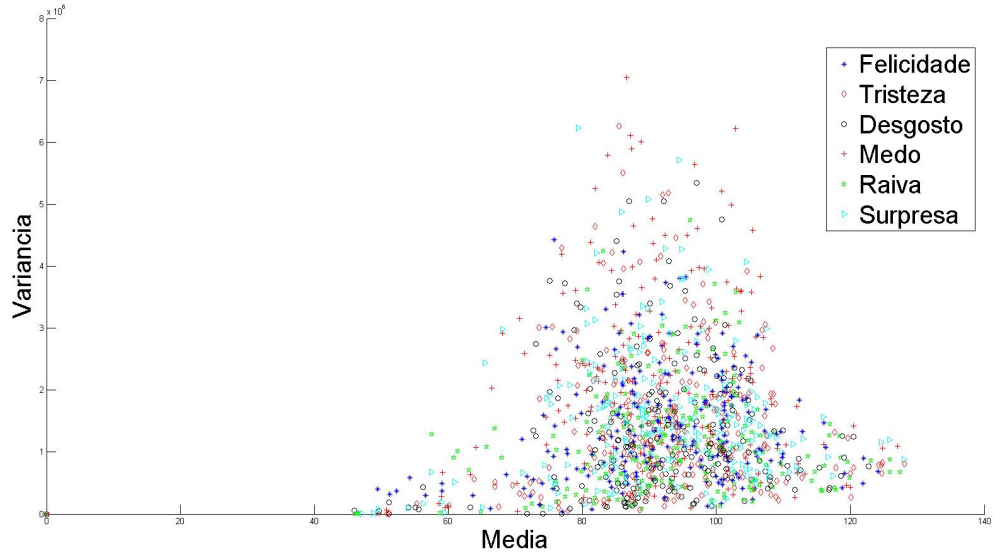


Figura 3.11: Descritores variância e média mostrados graficamente para seis emoções de todos os sujeitos.

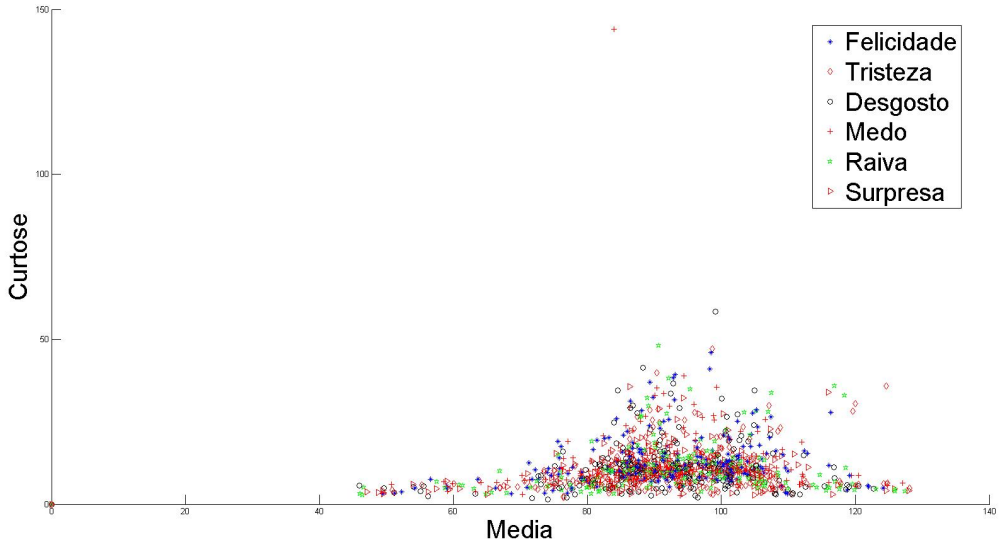


Figura 3.12: Descritores curtose e média mostrados graficamente para seis emoções de todos os sujeitos.

Foram feitos testes com duas RNAs: *Multilayer Perceptron* (MLP) e *Self-Organizing Maps* (SOM), ambos descritos em [16]. A RNA que obteve os melhores resultados foi a MLP, usando a rede *feedfoward*.

A Rede *Multilayer Perceptron* foi treinada usando o algoritmo *Levenberg-Marquardt*

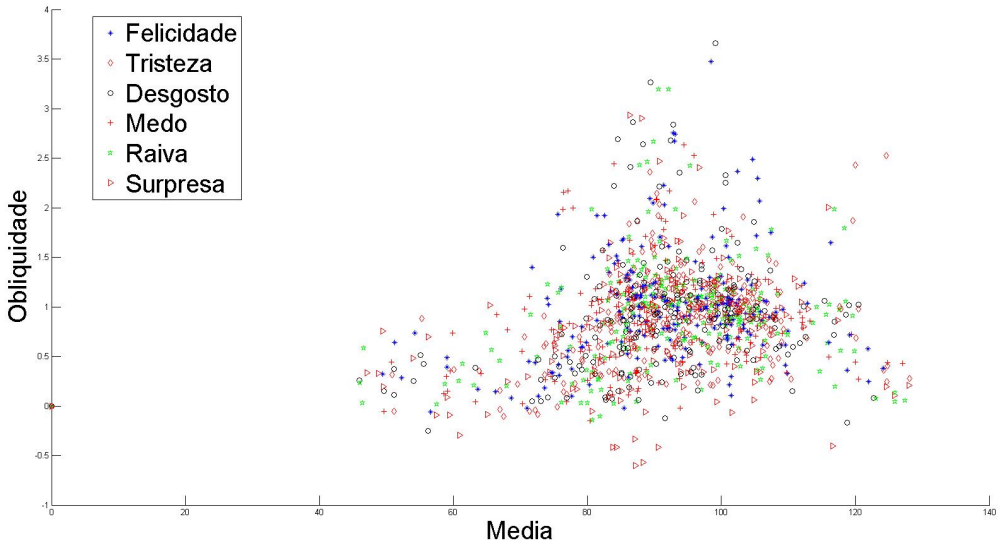


Figura 3.13: Descritores obliquidade e média mostrados graficamente para seis emoções de todos os sujeitos.

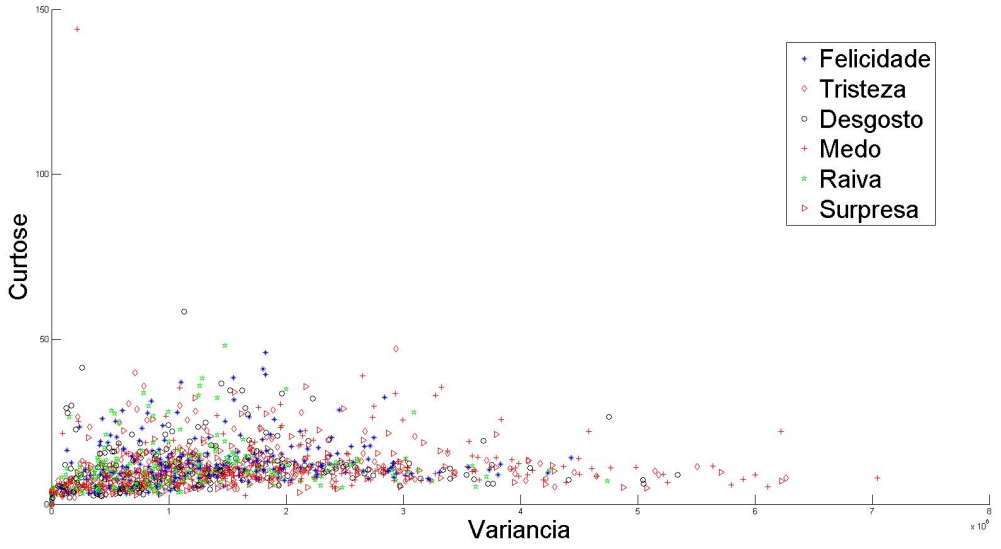


Figura 3.14: Descritores curtose e variância mostrados graficamente para seis emoções de todos os sujeitos.

[44]. As redes foram treinadas com diferentes quantidades de camadas e neurônios, dependendo do tamanho do conjunto de emoções que estavam sendo classificados.

Para a rede que classificou o conjunto de duas emoções, foram usadas cinco camadas, sendo uma de entrada, onde não é feito nenhum tipo de processamento computacional,

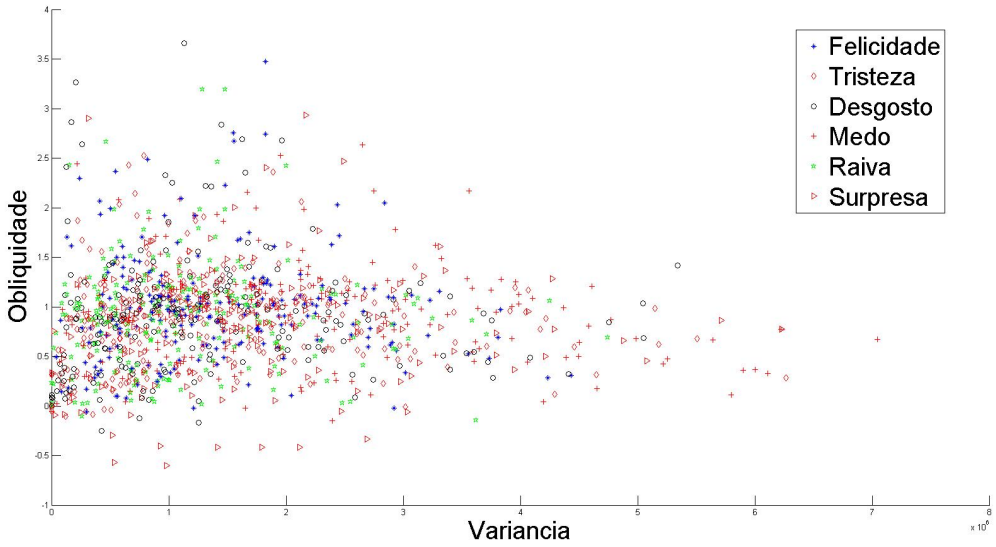


Figura 3.15: Descritores obliquidade e variância mostrados graficamente para seis emoções de todos os sujeitos.

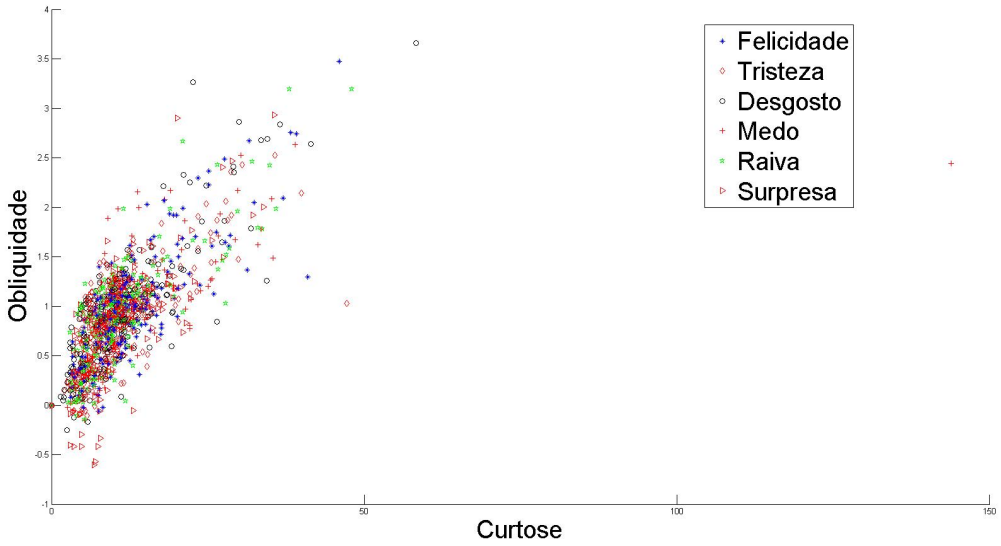


Figura 3.16: Descritores obliquidade e curtose mostrados graficamente para seis emoções de todos os sujeitos.

três camadas escondidas e uma camada de saída. A primeira camada simplesmente passa o vetor de entrada para o classificador. As camadas escondidas tem 17 neurônios, a primeira camada escondida tem uma função tangente hiperbólica de transferência, e as outras duas tem uma função linear de transferência. A camada de saída tem somente um

neurônio.

Com o conjunto de três emoções, por sua vez, a rede com melhores resultados foram obtidos com redes de quatro camadas, sendo novamente uma de entrada, duas escondidas e uma de saída. As camadas escondidas tem 129 neurônios, a primeira camada escondida usa uma função hiperbólica de transferência enquanto a segunda usa uma função linear de transferência.

Para o conjunto de quatro emoções, os melhores resultados tiveram quatro camadas, sendo novamente uma de entrada, duas escondidas e uma de saída. As camadas escondidas tinham 105 neurônios e a primeira camada escondida usava uma função hiperbólica de transferência enquanto a a segunda usava uma função linear de transferência.

A rede que classificou o conjunto de cinco emoções teve seis camadas, sendo uma de entrada, quatro escondidas e uma de saída. As camadas escondidas tinham 113 neurônios e primeira camada escondida usava uma função hiperbólica de transferência enquanto as outras três usavam uma função linear de transferência.

Para o conjunto completo de emoções, ou seja, seis emoções, a rede teve quatro camadas, sendo uma de entrada, duas escondidas e uma de saída. As camadas escondidas tinham 56 neurônios. A primeira camada escondida usava uma função hiperbólica de transferência enquanto a segunda usava uma função linear de transferência.

É sabido que redes neurais *feedfoward* com somente uma camada escondida podem aproximar qualquer função com um número finito de descontinuidades arbitrariamente bem [5]. No entanto, após testes exaustivos, foi percebido que em nossos experimentos os melhores resultados foram obtidos com pelo menos duas camadas escondidas.

Para treinar a rede, a saída foi treinada com vetores de 0s e 1s, onde o vetor continha 1 na emoção conhecida do vídeo e 0 nas outras.

Para classificar a emoção, o vetor de saída da rede é pego, e a posição de vetor com maior valor é dada como a emoção classificada.

3.2.5 Classificação usando a Rede Neural Artificial

Para classificar as emoções, videos variados com as emoções desmarcadas foram usados na RNA. O vetor de descritores foi criado e usado como entrada para a RNA. A rede irá usar esses valores para reconhecer a emoção.

A validação é feita comparando a emoção real do vídeo com a emoção reconhecida pela RNA, gerando uma matriz de confusão para determinar a efetividade da metodologia proposta.

Neste capítulo foi mostrada a metodologia do projeto, como o classificador funciona, mostrando todas as suas etapas para realizar a classificação. No próximo capítulo serão apresentados os testes feitos, bem como a base de dados usada para teste, e serão mostra-

dos os resultados do classificador, mostrando como este consegue classificar as emoções para cada conjunto de emoções.

Capítulo 4

Resultados

Neste capítulo serão apresentados os resultados obtidos em cada método usado na metodologia, sendo possível então identificar quais desses métodos servirão para serem utilizados como método de identificação de expressões faciais.

4.1 Base de dados

A base de dados usada na realização deste projeto é uma base pública de vídeos de emoção definida como *eNTERFACE'05*[22], composta por vídeos de 44 sujeitos distintos, onde cada sujeito possui gravações de vídeo representado 6 emoções distintas: raiva, felicidade, tristeza, medo, supresa e desgosto.

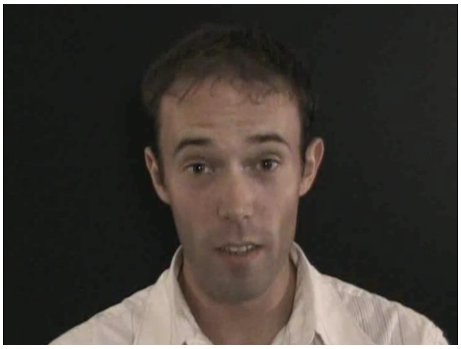
Para cada emoção temos 5 sentenças gravadas para cada sujeito, sendo que as sentenças são as mesmas para todos os sujeitos, ou seja, a sentença de número 1 da emoção raiva do sujeito 1 é a mesma que a sentença de número 1 da emoção raiva de todos os outros sujeitos. Dessa forma, temos então 44 sujeitos, com 6 emoções e 5 sentenças para cada emoção, totalizando 1320 vídeos de entrada.

As sentenças proferidas pelos sujeitos tem em média menos de 10 segundos, e o áudio é descartado no processamento, sendo usado somente o vídeo.

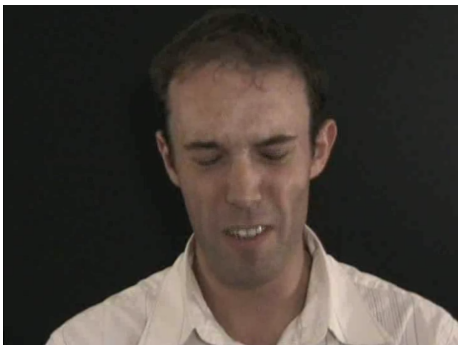
Na Figura 4.1 podemos ver *frames* de vídeos de 6 diferentes emoções de um mesmo sujeito.

4.2 Testes sem vídeos magnificados

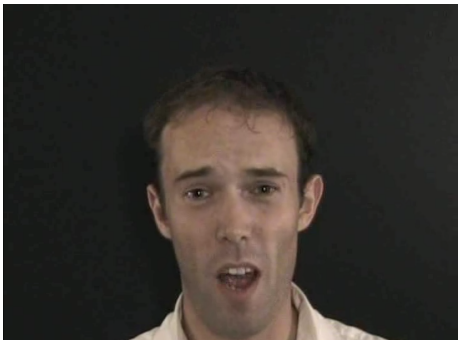
Nos testes, foram propostos cenários com diversos conjuntos de emoções, variando de duas a seis emoções e, exceto no caso no qual as seis emoções são usadas, variando também as emoções de entrada e validando os resultados para cada conjunto de emoção. Para



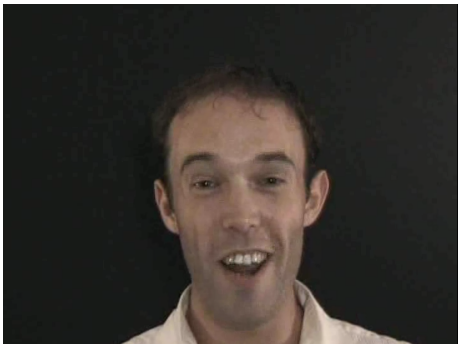
(a) Vídeo da emoção Raiva



(b) Vídeo da emoção Desgosto

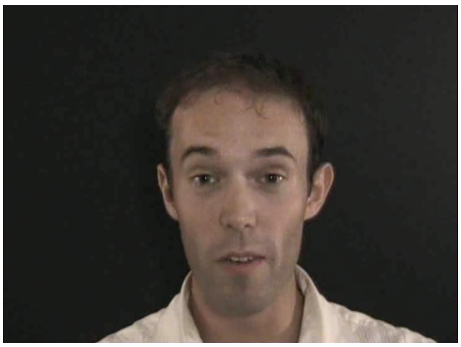


(c) Vídeo da emoção Medo



(d) Vídeo da emoção Felicidade

(e) Vídeo da emoção Tristeza



(f) Vídeo da emoção Surpresa

Figura 4.1: Exemplo de *frames* de vídeos das 6 emoções presentes na base

todos os cenários, uma estratégia de treinamento/avaliação de 70/30 (70% do conjunto foi usado no estágio de treinamento da RNA e 30% no estágio de avaliação) foi realizada.

Primeiramente, serão mostrados os testes com os vídeos não magnificados para futuramente mostrar a importância da magnificação na metodologia proposta

4.2.1 Conjunto de duas emoções

Foram feitos testes com conjuntos de duas classes de emoções, gerando matrizes de confusão, apresentadas nas Tabelas de 4.1 a 4.4 para testar a eficiência da rede em classificar vídeos não magnificados.

Emoção Detectada	Raiva	Medo
Raiva	0.406977	0.500000
Medo	0.593023	0.500000

Tabela 4.1: Matriz de confusão das emoções Raiva e Medo para o conjunto de duas emoções de todos os sujeitos.

Emoção Detectada	Desgosto	Felicidade
Desgosto	0.372093	0.290698
Felicidade	0.627907	0.709302

Tabela 4.2: Matriz de confusão das emoções Desgosto e Felicidade para o conjunto de duas emoções de todos os sujeitos.

Emoção Detectada	Raiva	Medo
Raiva	1	1
Medo	0	0

Tabela 4.3: Matriz de confusão das emoções Raiva e Medo para o conjunto de duas emoções do Sujeito 1.

Emoção Detectada	Desgosto	Felicidade
Desgosto	0.5	0
Felicidade	0.5	1

Tabela 4.4: Matriz de confusão das emoções Desgosto e Felicidade para o conjunto de duas emoções do Sujeito 1.

4.2.2 Conjunto de três emoções

Nesta subseção, mais uma emoção foi adicionada ao conjunto para verificar como a rede consegue classificar conjuntos de três emoções com os vídeos não magnificados, as matrizes de confusão geradas para validação estão apresentadas nas Tabelas de 4.5 a 4.8.

Emoção Detectada	Raiva	Medo	Tristeza
Raiva	0.511628	0.395349	0.348837
Medo	0.360465	0.360465	0.313953
Tristeza	0.127907	0.244186	0.337209

Tabela 4.5: Matriz de confusão das emoções Raiva, Medo e Tristeza para o conjunto de três emoções de todos os sujeitos.

Emoção Detectada	Raiva	Desgosto	Medo
Raiva	0.569767	0.418605	0.558140
Desgosto	0.197674	0.325581	0.186047
Medo	0.232558	0.255814	0.255814

Tabela 4.6: Matriz de confusão das emoções Raiva, Desgosto e Medo para o conjunto de três emoções de todos os sujeitos.

Emoção Detectada	Raiva	Medo	Tristeza
Raiva	1	0	0
Medo	0	1	0.5
Tristeza	0	0	0.5

Tabela 4.7: Matriz de confusão das emoções Raiva, Medo e Tristeza para o conjunto de três emoções do Sujeito 1.

Emoção Detectada	Raiva	Desgosto	Medo
Raiva	0.5	0	0
Desgosto	0.5	0.5	0
Medo	0	0.5	1

Tabela 4.8: Matriz de confusão das emoções Raiva, Desgosto e Medo para o conjunto de três emoções do Sujeito 1.

4.2.3 Conjunto de quatro emoções

Nesta subseção foram usados conjuntos de quatro emoções para validar a eficiência da rede em classificar vídeos não magnificados com quatro classes de emoções, as matrizes de confusão geradas para validar a eficiência podem ser vistas nas Tabelas de 4.9 à 4.12.

Emoção Detectada	Raiva	Medo	Felicidade	Desgosto
Raiva	0.453488	0.313953	0.220930	0.279070
Medo	0.220930	0.267442	0.313953	0.209302
Felicidade	0.139535	0.139535	0.197674	0.162791
Desgosto	0.186047	0.279070	0.267442	0.348837

Tabela 4.9: Matriz de confusão das emoções Raiva, Medo,Felicidade e Desgosto para o conjunto de quatro emoções de todos os sujeitos

Emoção Detectada	Medo	Desgosto	Surpresa	Tristeza
Medo	0.174419	0.174419	0.162791	0.186047
Desgosto	0.302326	0.313953	0.220930	0.325581
Surpresa	0.406977	0.453488	0.534884	0.360465
Tristeza	0.116279	0.058140	0.081395	0.127907

Tabela 4.10: Matriz de confusão das emoções Medo, Desgosto,Surpresa e Tristeza para o conjunto de quatro emoções de todos os sujeitos

Emoção Detectada	Raiva	Medo	Felicidade	Desgosto
Raiva	1	0	0	0
Medo	0	1	1	0.5
Felicidade	0	0	0	0
Desgosto	0	0	0	0.5

Tabela 4.11: Matriz de confusão das emoções Raiva, Medo,Felicidade e Desgosto para o conjunto de quatro emoções do Sujeito 1

Emoção Detectada	Medo	Desgosto	Surpresa	Tristeza
Medo	0	0	0	0
Desgosto	0	0	0	0
Surpresa	1	1	1	1
Tristeza	0	0	0	0

Tabela 4.12: Matriz de confusão das emoções Medo, Desgosto,Surpresa e Tristeza para o conjunto de quatro emoções do Sujeito 1

4.2.4 Conjunto de cinco emoções

Aqui, um conjunto com cinco classes de emoções é usado para verificar a eficiência da rede, as matrizes de confusão geradas para classificá-las podem ser encontradas nas Tabelas de 4.13 a 4.16.

Emoção Detectada	Raiva	Desgosto	Tristeza	Felicidade	Surpresa
Raiva	1	1	1	1	1
Desgosto	0	0	0	0	0
Tristeza	0	0	0	0	0
Felicidade	0	0	0	0	0
Surpresa	0	0	0	0	0

Tabela 4.13: Matriz de confusão das emoções Raiva, Desgosto, Tristeza , Felicidade e Surpresa para o conjunto de cinco emoções de todos os sujeitos.

Emoção Detectada	Medo	Tristeza	Raiva	Surpresa	Desgosto
Medo	0	0	0	0	0
Tristeza	0.988372	0.988372	0.988372	0.988372	0.976744
Raiva	0.011628	0.011628	0.011628	0.011628	0.023256
Surpresa	0	0	0	0	0
Desgosto	0	0	0	0	0

Tabela 4.14: Matriz de confusão das emoções Medo, Tristeza, Raiva, Surpresa e Desgosto para o conjunto de cinco emoções de todos os sujeitos.

Emoção Detectada	Raiva	Desgosto	Tristeza	Felicidade	Surpresa
Raiva	0.500000	0	0.500000	1.000000	1.000000
Desgosto	0.500000	1.000000	0.500000	0	0
Tristeza	0	0	0	0	0
Felicidade	0	0	0	0	0
Surpresa	0	0	0	0	0

Tabela 4.15: Matriz de confusão das emoções Raiva, Desgosto, Tristeza, Felicidade e Surpresa para o conjunto de cinco emoções do Sujeito 1.

4.2.5 Conjunto de seis emoções

Nesta seção serão feitos os testes com o conjunto completo de emoções, ou seja, as seis emoções, para testar como a rede classifica os vídeos sem magnificação. As matrizes de confusão geradas para validar os testes podem ser encontradas nas Tabelas 4.17 e 4.18.

Emoção Detectada	Medo	Tristeza	Raiva	Surpresa	Desgosto
Medo	1.000000	0	0	0	0
Tristeza	0	1.000000	0	0	0
Raiva	0	0	1.000000	0	0
Surpresa	0	0	0	0.500000	1.000000
Desgosto	0	0	0	0.500000	0

Tabela 4.16: Matriz de confusão das emoções Medo, Tristeza, Raiva, Surpresa e Desgosto para o conjunto de cinco emoções do Sujeito 1.

Emoção Detectada	Raiva	Desgosto	Medo	Felicidade	Tristeza	Surpresa
Raiva	0.2209	0.0930	0.1744	0.2093	0.0930	0.1395
Desgosto	0.1628	0.2907	0.2442	0.1860	0.3372	0.1977
Medo	0.2209	0.2326	0.1512	0.1744	0.2093	0.2558
Felicidade	0.0581	0.0349	0.0465	0.0581	0.0581	0.0465
Tristeza	0.2209	0.1395	0.2209	0.0930	0.1512	0.1279
Surpresa	0.2209	0.1395	0.2209	0.0930	0.1512	0.1279

Tabela 4.17: Matriz de confusão das emoções Raiva, Desgosto, Medo, Felicidade, Tristeza e Surpresa para o conjunto de seis emoções de todos os sujeitos

Emoção Detectada	Raiva	Desgosto	Medo	Felicidade	Tristeza	Surpresa
Raiva	0.5000	0	0	0	0	0
Desgosto	0	0.5000	0	0	0	0.5000
Medo	0	0	1.0000	0	0.5000	0
Felicidade	0	0	0	0.5000	0	0
Tristeza	0.5000	0	0	0	0.5000	0
Surpresa	0	0.5000	0	0.5000	0	0.5000

Tabela 4.18: Matriz de confusão das emoções Raiva, Desgosto, Medo, Felicidade, Tristeza e Surpresa para o conjunto de seis emoções do Sujeito 1

4.3 Testes com a magnificação ideal

Nesta seção serão mostrados os testes com a magnificação ideal, para mostras a importância da magnificaçào no processo de reconhecimento de emoções.

4.3.1 Conjunto de duas emoções

Neste cenário o conjunto de emoções é composto por duas classes de emoções, escolhidas aleatoriamente no conjunto de dados. Para avaliar a eficiência do algoritmo, a matriz de confusão desse cenário está descrita na Tabelas 4.19 a 4.22.

Emoção Detectada	Raiva	Medo
Raiva	0	0.5
Medo	1	0.5

Tabela 4.19: Matriz de confusão das emoções Raiva e Medo para o conjunto de duas emoções do Sujeito 1.

Emoção Detectada	Desgosto	Felicidade
Desgosto	0.5	0
Felicidade	0.5	1

Tabela 4.20: Matriz de confusão das emoções Desgosto e Felicidade para o conjunto de duas emoções do Sujeito 1.

Emoção Detectada	Raiva	Medo
Raiva	0.918605	0.872093
Medo	0.081395	0.127907

Tabela 4.21: Matriz de confusão das emoções Raiva e Medo para o conjunto de duas emoções de todos os sujeitos.

Emoção Detectada	Desgosto	Felicidade
Desgosto	0.697674	0.558140
Felicidade	0.302326	0.441860

Tabela 4.22: Matriz de confusão das emoções Desgosto e Felicidade para o conjunto de duas emoções de todos os sujeitos.

Apesar de apresentar uma situação simples, a matriz de confusão na Tabelas 4.19 a 4.20 mostraram que a RNA foi capaz de classificar as emoções sem dificuldade.

4.3.2 Conjunto de três emoções

Este cenário é composto por três classes de emoções, também obtidas aleatoriamente. A matriz de confusão deste cenário está descrita na Tabelas 4.23 a 4.26.

Emoção Detectada	Raiva	Medo	Tristeza
Raiva	1	0	0.5
Medo	0	1	0
Tristeza	0	0	0.5

Tabela 4.23: Matriz de confusão das emoções Raiva, Medo e Tristeza para o conjunto de três emoções do Sujeito 1.

Emoção Detectada	Raiva	Desgosto	Medo
Raiva	1	0	0
Desgosto	0	1	0
Medo	0	0	1

Tabela 4.24: Matriz de confusão das emoções Raiva, Desgosto e Medo para o conjunto de três emoções do Sujeito 1.

Emoção Detectada	Raiva	Medo	Tristeza
Raiva	0.441860	0.372093	0.337209
Medo	0.174419	0.244186	0.279070
Tristeza	0.383721	0.383721	0.383721

Tabela 4.25: Matriz de confusão das emoções Raiva, Medo e Tristeza para o conjunto de três emoções de todos os sujeitos.

Emoção Detectada	Raiva	Desgosto	Medo
Raiva	0.511628	0.395349	0.348837
Desgosto	0.209302	0.186047	0.255814
Medo	0.279070	0.418605	0.395349

Tabela 4.26: Matriz de confusão das emoções Raiva, Desgosto e Medo para o conjunto de três emoções de todos os sujeitos.

Para este cenário, houve um aumento na complexidade quando uma classe de emoções foi incluída. Analisando a matriz de confusão(Tabelas 4.23 a 4.26), há uma dificuldade para a RNA em classificar a emoção tristeza.

4.3.3 Conjunto de quatro emoções

Neste cenário, quatro classes de emoções são utilizadas, aumentando ainda mais a complexidade da RNA. Nas Tabelas 4.27 a 4.30 podemos ver as matrizes de confusões geradas para avaliar a eficiência da rede com um conjunto de quatro emoções.

Emoção Detectada	Raiva	Medo	Felicidade	Desgosto
Raiva	0.337209	0.162791	0.162791	0.209302
Medo	0.348837	0.372093	0.209302	0.348837
Felicidade	0.232558	0.337209	0.523256	0.348837
Desgosto	0.081395	0.127907	0.104651	0.093023

Tabela 4.27: Matriz de confusão das emoções Raiva, Medo,Felicidade e Desgosto para o conjunto de quatro emoções de todos os sujeitos

Emoção Detectada	Medo	Desgosto	Surpresa	Tristeza
Medo	0.162791	0.093023	0.162791	0.151163
Desgosto	0.255814	0.302326	0.197674	0.348837
Surpresa	0.395349	0.395349	0.348837	0.220930
Tristeza	0.186047	0.209302	0.290698	0.279070

Tabela 4.28: Matriz de confusão das emoções Medo, Desgosto,Surpresa e Tristeza para o conjunto de quatro emoções de todos os sujeitos

Emoção Detectada	Raiva	Medo	Felicidade	Desgosto
Raiva	1	0	0	0.5
Medo	0	1	0	0.5
Felicidade	0	0	1	0
Desgosto	0	0	0	0

Tabela 4.29: Matriz de confusão das emoções Raiva, Medo,Felicidade e Desgosto para o conjunto de quatro emoções do Sujeito 1

Emoção Detectada	Medo	Desgosto	Surpresa	Tristeza
Medo	0.5	0	0	0
Desgosto	0	0.5	0	0.5
Surpresa	0.5	0.5	1	0
Tristeza	0	0	0	0.5

Tabela 4.30: Matriz de confusão das emoções Medo, Desgosto,Surpresa e Tristeza para o conjunto de quatro emoções do Sujeito 1

Analisando as matrizes de confusão podemos ver que a rede está começando a perder a sua eficiência na classificação das emoções. Na Tabela 4.29 é possível ver que as emoções Raiva e Felicidade foram bem classificadas, enquanto as emoções Desgosto e Medo não. No entanto a Tabela 4.30 mostra que para alguns conjuntos de emoções a RNA já não classifica bem nenhuma emoção deste conjunto.

Analisando a Tabela 4.28 com a Tabela 4.30 é possível notar que aumentando a quantidade de sujeitos a eficiência da rede diminui. Porém, ao analisar a Tabela 4.27 com a Tabela 4.29, é perceptível que o teste feito com todos os sujeitos tem melhores resultados que o teste feito com somente um sujeito, isto porque para alguns sujeitos a rede conseguiu classificar melhor do que outros, e o sujeito 1 foi um dos sujeitos em que a rede não conseguiu classificar bem.

4.3.4 Conjunto de cinco emoções

Neste cenário, são usadas cinco classes de emoções. As matrizes de confusão obtidas (Tabelas 4.31 a 4.34) mostraram que as capacidades discriminativas da RNA, quando usada a configuração descrita, teve sua complexidade aumentada, após incluir mais classes de emoções. No entanto a RNA, não conseguiu classificar bem a emoção raiva.

Emoção Detectada	Raiva	Desgosto	Tristeza	Felicidade	Surpresa
Raiva	1	1	1	1	1
Desgosto	0	0	0	0	0
Tristeza	0	0	0	0	0
Felicidade	0	0	0	0	0
Surpresa	0	0	0	0	0

Tabela 4.31: Matriz de confusão das emoções Raiva, Desgosto, Tristeza, Felicidade e Surpresa para o conjunto de cinco emoções do Sujeito 1.

Emoção Detectada	Medo	Tristeza	Raiva	Surpresa	Desgosto
Medo	1	0	0	0	0.5
Tristeza	0	0.5	0	0	0
Raiva	0	0	1	0	0
Surpresa	0	0	0	0.5	0
Desgosto	0	0.5	0	0.5	0.5

Tabela 4.32: Matriz de confusão das emoções Medo, Tristeza, Raiva, Surpresa e Desgosto para o conjunto de cinco emoções do Sujeito 1.

É perceptível nesse cenário que com um conjunto de cinco classes de emoções, a RNA não conseguiu classificar as emoções, nem mesmo a classificação com somente um

Emoção Detectada	Raiva	Desgosto	Tristeza	Felicidade	Surpresa
Raiva	0.279070	0.232558	0.220930	0.127907	0.244186
Desgosto	0.151163	0.255814	0.325581	0.162791	0.162791
Tristeza	0.174419	0.081395	0.209302	0.081395	0.139535
Felicidade	0.104651	0.186047	0.069767	0.290698	0.127907
Surpresa	0.290698	0.244186	0.174419	0.337209	0.325581

Tabela 4.33: Matriz de confusão das emoções Raiva, Desgosto, Tristeza , Felicidade e Surpresa para o conjunto de cinco emoções de todos os sujeitos.

Emoção Detectada	Medo	Tristeza	Raiva	Surpresa	Desgosto
Medo	0.081395	0.046512	0.081395	0.127907	0.081395
Tristeza	0	0	0	0	0
Raiva	0	0	0	0	0
Surpresa	0.348837	0.209302	0.325581	0.348837	0.395349
Desgosto	0.569767	0.744186	0.593023	0.523256	0.523256

Tabela 4.34: Matriz de confusão das emoções Medo, Tristeza, Raiva, Surpresa e Desgosto para o conjunto de cinco emoções de todos os sujeitos.

sujeito nem a classificação feita com todos os sujeitos. A principal causa desta falha na classificação proposta se deve ao fato de que o processo de magnificação de vídeo, devido ao grande movimento e baixa taxa de *frames* por segundo, gerou uma alta quantidade de ruído, o que faz com que o vetor de descritores proposto não consiga uma classificação apropriada.

4.3.5 Conjunto de seis emoções

Neste cenário, foram usadas todas as seis classes de emoções do conjunto de dados. Os resultados obtidos estão descritos nas matrizes de confusão nas Tabelas 4.35 e 4.36.

Emoção Detectada	Raiva	Desgosto	Medo	Felicidade	Tristeza	Surpresa
Raiva	0.2326	0.0930	0.2209	0.2093	0.1977	0.1860
Desgosto	0.1512	0.2442	0.1744	0.1047	0.2326	0.1628
Medo	0.1512	0.0814	0.1163	0.1047	0.2326	0.1047
Felicidade	0.3023	0.3605	0.2326	0.3953	0.2093	0.3953
Tristeza	0.1279	0.1163	0.1047	0.0349	0.0465	0.0698
Surpresa	0.0349	0.1047	0.1512	0.1512	0.0814	0.0814

Tabela 4.35: Matriz de confusão das emoções Raiva, Desgosto, Medo, Felicidade, Tristeza e Surpresa para o conjunto de seis emoções de todos os sujeitos

Emoção Detectada	Raiva	Desgosto	Medo	Felicidade	Tristeza	Surpresa
Raiva	0.5000	0	0	0	0	0
Desgosto	0.5000	0.5000	0	0	0.5000	0
Medo	0	0.5000	0.5000	0	0	0
Felicidade	0	0	0	0.5000	0.5000	0.5000
Tristeza	0	0	0	0	0	0
Surpresa	0	0	0.5000	0.5000	0	0.5000

Tabela 4.36: Matriz de confusão das emoções Raiva, Desgosto, Medo, Felicidade, Tristeza e Surpresa para o conjunto de seis emoções do Sujeito 1

Neste cenário, é possível perceber que as classificações com seis emoções não foram satisfatórias, e que como os outros cenários, a classificação feita com apenas um sujeito foram melhores do que a classificação feita com todos os sujeitos.

Comparando as matrizes de confusão de todos os casos dos vídeos não magnificados com os vídeos magnificados é perceptível que a magnificação melhorou muito a eficácia da classificação da RNA, justificando então a utilização da MVE na classificação de emoções.

Capítulo 5

Conclusão

Neste projeto, é apresentada uma estratégia inicial para reconhecer diferentes emoções em um vídeo usando MVE e RNAs. Baseado nos resultados apresentados, é possível confirmar que a classificação indica uma grande possibilidade de usar a metodologia proposta em um grupo restrito de emoções.

É possível concluir primeiramente que o algoritmo consegue classificar bem com um pequeno conjunto de emoções, mas perde sua eficiência a medida que o conjunto de emoções aumenta.

É notável que algumas emoções possuem comportamentos similares, como Raiva e Desgosto, confundindo a classificação. Da mesma forma, existem também emoções que são completamente diferentes, como a Raiva e a Surpresa, sendo portanto mais fáceis de serem classificadas.

Comparando as matrizes de confusão de todas as pessoas analisadas, é notável que o algoritmo trabalha melhor com vídeos da mesma pessoa do que com vídeos de diferentes pessoas juntas. Nos testes iniciais, o comportamento foi esperado, já que pessoas diferentes expressam suas emoções de maneiras diferentes, ainda que estejam dizendo a mesma frase.

Mesmo com os problemas listados acima, é possível dizer que o algoritmo funciona bem para classificar emoções que possuem pouco grau de similaridade.

5.1 Trabalhos Futuros

Trabalhos futuros irão incluir um aperfeiçoamento da implementação, usando um grupo maior de emoções e mais pessoas a serem testadas, e para evitar problemas relacionados com as RNAs usuais, outros candidatos para classificadores podem ser testados, como redes neurais convolutivas.

No aperfeiçoamento da implementação será possível incluir tentativas de se classificar emoções de diferentes sujeitos e uma melhoria para classificar emoções similares, possivelmente usando novos descritores para isto.

Referências

[1] M.Y. Alva, M. Nachamai, and J. Paulose. A comprehensive survey on features and methods for speech emotion detection. In *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on*, pages 1–6, March 2015. 4

[2] E.M. Bouhabba, A.A. Shafie, and R. Akmeliawati. Support vector machine for face emotion detection on real time basis. In *Mechatronics (ICOM), 2011 4th International Conference On*, pages 1–6, May 2011. 4

[3] Mark Cox CI2CV, Computer Vision Lab, 2013. 2

[4] Domiciano Correa Marques da Silva, Índice de refração e a dispersão da luz. x, 17

[5] Howard Demuth, Mark Beale, Howard Demuth, and Mark Beale. Neural network toolbox for use with matlab, 1993. 31

[6] Mark Hudson Beale Martin T. Hagan Howard B. Demuth. *Neural Network Toolbox^{TM} User's Guide*. MathWorks, The MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098, 2015. x, 11, 13

[7] P. Ekman, W.V. Friesen, and J.C. Hager. The facial action coding system. In *Research Nexus eBook*, 2002. 4

[8] Speech Emotion Recognition EMOSpeech, 2012. 3

[9] Emotions drive spending emotient, 2015. 4

[10] M. Fairhurst, M. Erbilek, and Cheng Li. Enhancing the forensic value of handwriting using emotion prediction. In *Biometrics and Forensics (IWBF), 2014 International Workshop on*, pages 1–6, March 2014. 1

[11] Yuchao Fan, Mingxing Xu, Zhiyong Wu, and Lianhong Cai. Automatic emotion variation detection using multi-scaled sliding window. In *Orange Technologies (ICOT), 2014 IEEE International Conference on*, pages 232–236, Sept 2014. 5

[12] Noureddine CHERABIT Fatma zohra CHELALI and Amar DJERAD. Face recognition system using skin detection in rgb and ycbcr color space. *laboratory,Faculty of Electronics engineering and computer science, National School of Technology ENST, Rouiba*, pages 1 – 7, mar 2015. 18

[13] Master Student Francesco Bonadiman, 2015. 4

[14] Inc. Free Software Foundation, 1989, 1991. x, 12, 21

[15] Artificial Neural Network Glosser.ca, 2013. x, 13

[16] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998. 11, 28

[17] D. Jagadiswary, G. Appasami, and S. Rajesh. Eye features normalization and face emotion detection for human face recognition. In *Electronics, Communication and Computing Technologies (ICECCT), 2011 International Conference on*, pages 64–68, Sept 2011. 4

[18] E. Leon, G. Clarke, F. Sepulveda, and V. Callaghan. Real-time physiological emotion detection mechanisms: Effects of exercise and affect intensity. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 4719–4722, 2005. 1

[19] YCbCr LionDoc, 20012. x, 18

[20] Nguyen M.K. Liu Y., Sourina O. Real-time eeg-based emotion recognition and its applications. *Special Issue on Cyberworlds*, 2011. 4

[21] Yoshinari Kameda Takehisa Onisawa Maria Alejandra Quiros-Ramirez, Senya Polikovsky. Towards developing robust multimodal databases for emotion analysis. *Humanistic Systems Laboratory, Computer Vision and Image Laboratory*. 2

[22] O. Martin, I. Kotsia, B. Macq, and I. Pitas. The enterface' 05 audio-visual emotion database. In *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on*, pages 8–8, April 2006. 33

[23] Nancy Al Haddad Mohamad Fadel Michel Owayjan, Ahmad Kashour and Ghinwa Al Souki. The design and development of a lie detection system using facial micro-expressions. *Department of Computer and Communications Engineering*. 2

[24] YCbCr Mike1024, 2006. x, 19

[25] Robinson P. Ntombikayise B. Multimodal affect recognition in intelligent tutoring systems. *Proceedings of the 4th international conference on Affective computing and intelligent interaction*, 2011. 3

[26] Byung-Hun Oh and Kwang-Seok Hong. A study on facial components detection method for face-based emotion recognition. In *Audio, Language and Image Processing (ICALIP), 2014 International Conference on*, pages 256–259, July 2014. 5

[27] Maja Pantic and Leon J. M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. In *Proceedings of the IEEE*, pages 1370–1390, 2003. 4

[28] Perkusich A. Rached T.S. Emotion recognition based on brain-computer interface systems. *Brain-Computer Interface Systems*, 2013. 3

[29] Richard E. Woods Rafael C. Gonzalez. *Digital Image Processing(3rd Edition)*. Pearson International Education, 2006. 16

[30] Dolly Reney and Neeta Tripathi. An efficient method to face and emotion detection. In *Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on*, pages 493–497, April 2015. 4

[31] Gregory Hager Rizwan Chaudhry, Avinash Ravichandran and Rene Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. *Center for Imaging Science, Johns Hopkins University.* 2

[32] David Sanchez-Mendoza, David Masip, and Agata Lapedriza. Emotion recognition from mid-level features. *Pattern Recognition Letters*, 67, Part 1:66 – 74, 2015. Cognitive Systems for Knowledge Discovery. 1

[33] creepy taste of the future of wearable computers Sebastian Anthony, Real-time emotion detection with Google Glass: An awesome, 2014. 3

[34] Bo-Hao Su, Ping-Wen Fu, Po-Chuan Lin, Po-Yi Shih, Yuh-Chung Lin, Jhing-Fa Wang, and An-Chao Tsai. A spoken dialogue system with situation and emotion detection based on anthropomorphic learning for warming healthcare d. In *Orange Technologies (ICOT), 2014 IEEE International Conference on*, pages 133–136, Sept 2014. 4

[35] Ching-Chih Tsai, You-Zhu Chen, and Ching-Wen Liao. Interactive emotion recognition using support vector machine for human-robot interaction. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 407–412, Oct 2009. 1

[36] Paul Viola and Michael J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004. 22

[37] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. Phase-based video motion processing. *ACM Trans. Graph. (Proceedings SIGGRAPH 2013)*, 32(4), 2013. x, 8

[38] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (Proc. SIGGRAPH 2012)*, 31(4), 2012. x, 6, 7, 8, 9, 10, 12, 20, 23

[39] Changsheng Xu Qi Tian Hanqing Lu Xiaofeng Tong, Lingyu Duan. Local motion analysis and its application in video based swimming style recognition. *National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing China 100080.* 2

[40] and Randall Davis Yale Song, Louis-Philippe Morency. Learning a sparse codebook of facial and body microexpressions for emotion recognition. *MIT CSAIL , and USC ICT.* 2

[41] Yee Ying Yick, Luciano Grüdtner Buratto, and Alexandre Schaefer. The effects of negative emotion on encoding-related neural activity predicting item and source recognition. *Neuropsychologia*, 73:48 – 59, 2015. 1

[42] Park K.-S. Yoon W.-J. A study of emotion recognition and its applications. *4th International Conference, MDAI, Springer*, 2007. 3

[43] Park K.-S. Yoon W.-J. A study of speech emotion recognition and its application to mobile services. *4th International Conference, UIC, Springer*, 2007. 4

[44] Hao Yu and B. M. Wilamowski. Levenberg–marquardt training. In *Industrial Electronics Handbook, 2nd Edition*, volume 5, pages 12–1. CRC Press, 2011. 29